

The year of utilitarian science?

The world's governments are remarkably at one in believing that they must win value from the research they sponsor, but they too often overlook the difficulties in their way.

THE convention that a new year should prompt questions about the past and about the future is a good convention. This new year (on the Gregorian calendar) is especially apposite for those with an interest in the future of the scientific enterprise. Now, more than five years after the collapse of the Berlin Wall, it is clear that the much anticipated 'peace dividend' will not be used to usher in another golden age of discovery. In retrospect, we should not be surprised. The Cold War was only one of many outstanding problems at the end of 1989. Its end has merely meant that resources have been shifted in other directions — capital investment on a huge scale in eastern Asia, slightly more generous assistance for economically deprived countries, for example.

Yet it is a genuine surprise that the response of governments should have been so uniform, and so swift. At the beginning of 1994, a news review pointed to some of the political explanations (see *Nature* 367, 5–9; 1994): in many places, the end of the Cold War was recognized as a signal that military competition had been replaced by economic competition. This year's version of that review (see pages 5–9) details the ways in which governments are seeking to win direct economic benefits from their support of science. That individual governments should take that line is not remarkable. The surprise is the generality of the cry for utilitarian science.

What will the new utilitarian science be like? One trend is already plain: it will increasingly fall to industrial companies (rather than to public agencies) to support research in which they have a direct interest. Within limits, that makes sense, and can work well. In both Japan and Switzerland, most research is already privately supported. Now, the principles of international free trade require no less.

Record

Company-based research has a good record in the past few decades. The remarkable development of the computer industry in the past 20 years, first in the United States and then Japan, has been almost entirely financed by the companies likely to benefit, or hoping soon to do so. The international pharmaceutical industry has an even longer record of successful self-reliance. So why should not governments let companies get on with what are their own affairs? For practical purposes, that is what the government of Japan has done since the Second World War, with enviable success. Nobody should be surprised when the new Republican Congress in the United States begins singing the same

refrain, perhaps in the guise of 'getting the government off the people's back'.

Snags

There are several snags. The most obvious is that even successful companies have a continuing need for able people trained in research who are best equipped for the practice of innovation by taking research degrees at universities. So governments find that even industrial companies able and willing to finance their own research clamour for trained researchers. In the emerging economies of Asia, new universities are now scrambling to found PhD courses from which, they hope, generations of innovators will emerge. The difficulty they will find is that which has become apparent in Japan, where companies have traditionally shouldered much of the burden of education in research: the research labour force is insufficiently mobile, and there is an insufficient flow of radical ideas from a vigorous community active in basic research.

That is the case for asking that governments, as a national obligation, should continue to support their research universities as institutions as well as a hard core of research students. But even that principle is, in some places, under threat from the doctrines of utilitarian science. For example, the endless squabbling in the United States about the proper level of 'indirect costs' (overhead payments) attributable to research grants is a sign of backsliding that will most seriously damage the private universities, many of which are among the most productive of important innovation. In Germany, the devolution of responsibility for higher education to the *Länder* governments may also inhibit a proper nurturing of the research universities, the excellent work of the *Deutscher Forschungsgemeinschaft* and the *Max-Planck Gesellschaft* notwithstanding. Britain, after more than a decade of agonizing, seems to have found a solution that, on paper, is more nearly correct: research universities, designated as such on the basis of periodic assessments, are paid extra running costs and then compete for research grants that also carry a modest overhead. We shall have to see how well the system works.

The other and more serious stumbling-block in the new utilitarian regime is that of what constitutes appropriate research for research universities to pursue. The present fashion is that academic research should be 'relevant' research, angled more or less directly at the solution of practical problems. In principle, there is nothing wrong with that.



The essence of a training in research is that a student should confront that unique dilemma of winning an understanding of the natural world that nobody has previously found; almost any problem will serve the purpose, but the techniques employed are also relevant to future employment. But the needs of relevance create problems for university departments, not least those of continuity of intellectual interest. Academic departments are most productive both of discovery and of trained people when they are centred on long-term research goals, and when students in their research training contribute to the attainment of those goals. Those who preach the benefits of utilitarian science too often overlook that well-attested principle, jumping to the false conclusion that what they demean as 'curiosity-driven' research is by definition irrelevant.

The other widely neglected principle is that academic research establishments must, to be effective, be impartial sources of knowledge and skill for all their potential customers. In the abstract, it will readily be agreed that a university even partly supported with public funds should not be in the pocket of a single company, to the exclusion of its competitors. Properly, the value of its expertise is at least a national asset, often an international one. Yet the demand for relevance often translates into pressure for universities and departments to form research partnerships with industrial companies, whose interests rightly exclude those of their competitors. That does not imply that every research deal struck between a company and a university is suspect; many of them are enlightened ways in which research basic to some general interest of the company can be adequately supported. But there are dangers in these arrangements more often obvious to outsiders than to those who make them. Research universities will need, in the years ahead, to be stout in their resistance to pressure from cost-conscious administrators to form more partnerships of this kind.

There remains the issue of the people who work in research universities and other establishments with comparable functions. Utilitarian science will put a raft of unfamiliar ideas into their heads. If utilitarian science is that much more relevant to wealth creation and even to the solution of global problems such as climatic change, should they not be rewarded differently, and more generously, than at present? This question will arise with particular force in the minds of the unsung army of the world's researchers, the postdoctoral fellows who are usually the ones who switch off the laboratory lights at night. If, as the planners tell us, utilitarian science will be generally more useful science, is it not right that its practitioners should share in the benefits, especially if they are required (as they should be) to preserve their impartiality from external exclusive entanglements?

In short, it seems probable that the implementation of utilitarian science will prove to be more difficult than its advocates believe. Many of the outstanding problems are issues of principle that will not be quickly solved, but there will also be some hesitation when it turns out that utilitarian science is not cheaper, but more expensive, than the old variety. To be sure, to the extent that many governments now seem willing to let their national companies sink or swim by

the degree to which they support imaginative innovation policies with their own resources, the rest of the world may yet follow Japan and Switzerland towards prosperity. But there is also a danger, yet to be assessed, that the new regime, if too quickly and insensitively established, will persuade many active and able researchers that they had better do something else instead. The effect on young people, already in the West uncomfortably unwilling to follow technical careers, could be even more damaging. □

Nature and Le Monde

This journal plans an exercise in science popularization with a leading French newspaper.

SOCIAL occasions are often a great embarrassment for members of the staff of this journal, who will invariably be told by non-scientists they meet that they have, "of course", heard of *Nature*, but that they would in no circumstances expect to read it or understand a word if somebody should thrust a copy into their hands. That is a great source of frustration to people who correctly believe that *Nature* carries many accounts of developments in the understanding of the natural world that rank in interest and importance with the other contemporary peaks of intellectual life.

That frustration is one reason why *Nature* has gladly entered into an arrangement with the great French daily newspaper *Le Monde* to produce what is hoped will be a interesting exercise in popularization. The immediate goal is that the science staff of the newspaper and that of *Nature* will work together on the production of single pages of the newspaper, each dealing with a topic in science (not necessarily prompted by research articles appearing in this journal), that will appear every other week (beginning next week). The hope is that, by a blend of original reporting and the exposition of technical matters, it will be possible to produce something distinctive and worthwhile.

But why an arrangement with a French newspaper when the world is full of English-language sheets? There are two explanations. For one thing, French people who are not scientists retain an intellectual interest in the nature of discovery that probably surpasses that elsewhere in Europe. Then, with the general acceptance of English as the language of the scientific journals, there is a danger that science will come to seem an Anglo-Saxon pursuit. There are the strongest reasons for demonstrating that not to be the case.

But journalists on other newspapers than *Le Monde* should not regard this arrangement as a threat to the relations (hopefully good) they now enjoy with *Nature*. The means now used for alerting them to the contents of each week's issue will continue. Nor need journalists elsewhere fear that embargo dates enforced for them will be broken for the benefit of *Le Monde*. What is exclusive in the new arrangement is simply the hope that careful discussion among two groups of science journalists with a common interest in intelligent popularization will yield something to be proud of in a difficult field. □

US National Cancer Institute chief quits

Washington. Sam Broder, director of the National Cancer Institute (NCI) since 1989, is to leave the institute in April to become research chief at IVAX, a pharmaceutical company based in Miami, Florida. Broder says that he wants to leave the NCI, where he has worked for 22 years, while he is still young enough to pursue a career in the private sector. He is 49.

He denied reports that his departure is the result of disagreements with Harold Varmus, director of the National Institutes of Health (NIH), the NCI's parent organization, about a planned restructuring of the institute. "We get along very well," Broder said. "We don't agree on everything — but if everyone always agreed, it would be a sign of a dysfunctional system. You don't want yes-men in high positions at NIH, and Dr Varmus rather enjoys that give and take."

Varmus issued a statement praising Broder's record in research and management at the institute, and thanking him for "advice to me during my first years as director of NIH". With an annual budget of \$2.2 billion, the NCI is by far the largest component of the NIH, and its director has considerable autonomy.

Broder said his departure was also prompted by frustration with the slow

pace of decision-making in government. Over the years, he said, "the bureaucracy has gotten worse, and my tolerance of it has gotten less".

He expressed satisfaction with his achievements during his time as NCI's director. "We promoted the concept that NCI is an organic whole, consisting of basic research, clinical research and the cancer centres. I have strengthened all three in the last six years. "We also won the largest dollar increase in our history in 1992. That was, in part, a vote of confidence in my leadership."

At IVAX — a \$600-million company known best for its generic drugs — Broder will become vice president for research and development.

Broder is widely liked and respected, even by his critics among cancer activists and in Congress. Some now fear a leadership vacuum at NCI, where the deputy directorship and several other senior positions are already vacant.

A successor will be appointed by President Bill Clinton, presumably in consultation with Varmus and Donna Shalala, the health secretary. The post does not require Senate confirmation, but the administration may still be hard-pressed to fill it before Broder leaves in April.

Colin Macilwain

Nirex to appeal after permission refused for 'rock lab' at Sellafield

London. Cumbria County Council last month refused planning permission for a £120-million rock characterization facility (RCF), that the company UK Nirex Ltd wants to build at Sellafield in Cumbria, in northwest England, near the planned site for the United Kingdom's first underground repository for nuclear waste. Nirex says it will appeal against the decision.

Nirex's plans for the RCF — which would be used to study groundwater flow — are backed by the Royal Society. The society recently recommended that the rock laboratory should be built "as soon as is practicable", provided that sufficient information was obtained before the RCF disturbs the area's hydrogeological characteristics (see *Nature* 372, 309; 1994).

But the decision by Nirex to appeal will almost certainly lead to a public inquiry into the application. The council had already requested a comprehensive inquiry in a letter to John Gummer, the secretary of state for the environment, last August, shortly after it received the application for planning permission — the RCF, it told Gummer "raises matters of more than local importance".

The council voted overwhelmingly to reject Nirex's application, at a meeting on 20 December, but Gummer had in any case already instructed the council "not to grant permission on the application without special authorization". Gummer said he needed more time in which to decide whether to call a public inquiry.

Because the council had rejected the application early in the meeting, it did not vote on another motion to defer the application until the government had completed reviews of the nuclear industry and radioactive waste, and until Nirex had obtained more data from existing boreholes around the proposed repository zone. Some observers consider that this voting procedure was abnormal, and that the motions should have been proposed, amended and debated simultaneously.

Nirex says it will appeal as soon as it "gets the relevant papers together". It added that the public inquiry, which will take several months to organize, will probably be able to take into account the conclusions of the government reviews now under way. "The whole thing now will be very timely."

Some observers argue that the application was bound to be rejected because Nirex made it too soon. But Nirex argues that it needs to push ahead with hydrogeological investigations, and is continually publishing data from existing boreholes.

Maggie Verrall

Newton adds gravitas to new library

London. This 12-foot high statue of Sir Isaac Newton, plotting the immensity of the Universe with a pair of dividers, will be the first sight to greet visitors entering the grounds of the new British Library in St Pancras in London, when it opens in late 1997.

The statue, which is inspired by William Blake's famous painting of Isaac Newton, will stand on a 12-foot plinth in the Piazza, framed by the main entrance gates. Scottish sculptor Eduardo Paolozzi, who is seen with his statue as it nears completion, was commissioned to make the piece after the architect of the new building, Colin StJohn Wilson, saw a scaled-down model in Paolozzi's studio.

Delays in the construction of the building and problems with the project running over budget have attracted criticism. Stephen Dorrell, the secretary of state for national heritage, announced after the budget in November last year that an extra £46 million would be made available to complete construction by the end of 1996. Previous estimates put the total cost for the building at £450 million.



Call for plutonium shipment to be postponed

Washington. Greenpeace International, the Washington-based Nuclear Control Institute and the Tokyo-based Citizens Nuclear Information Center, last month released a report supporting their protests against a forthcoming shipment of reprocessed nuclear waste from France to Japan.

The first commercial-scale shipment of plutonium extracted in France from spent fuel rods was returned to Japan in early 1993. The new shipment would be the first of about 30 over the next 15 years to transport additional waste from fuel rods.

But the report, by Edwin S. Lyman, a physicist at Princeton University in New Jersey, argues that the shipment should be postponed until its safety has been more fully assessed. In particular, Lyman claims that current international safety standards for containing vitrified high-level waste may not be stringent enough. The tests of shipping casks on which these standards are based do not take sufficiently into account extreme conditions of fire, collision and immersion, he argues.

Lyman also claims that the radioactive leakage rates permitted by the Japanese government may also be too high. He says that the maximum radiation doses to which those close to casks might be

exposed have been determined using limited data and rough estimates of cask leak rates, and the relation between radionuclide emission rates and the amount of radiation leaking from the casks.

The report also claims that company literature gives a maximum permitted storage and transport temperature of 510 °C. Lyman argues that this temperature could damage the casks and their rubber seals, and is close to the 600 °C threshold where the glass matrix around the waste could begin to undergo "undesirable changes".

British Nuclear Fuels Limited (BNFL) — which owns 60 per cent of Pacific Nuclear Transport Limited, the company handling the shipment — says the report seems to be based on "dubious science". It also claims that the shipment would be carried out at 200–250 °C and not 510 °C. The report, it says, "raises no new issues".

Speaking at a Washington press conference where the report was released, US congressmen and public-interest groups also criticized secrecy over the timing and route of the shipment's two-month voyage. This would make it difficult, they argued, to arrange contingency plans in the event of an emergency.

Julie Wakefield

Indian scientists' missed opportunity

New Delhi. **Research by the historian S. Ramaseshan, who is also editor of the Indian journal *Current Science*, has shown that disagreements among Indian scientists thwarted efforts by political leaders during the run-up to independence in 1947 to merge the three existing science academies into a "United Academy of Sciences of India", which would have provided the new government with a single source of advice on scientific issues.**

Ramaseshan uncovered the information in the archives of the Indian Academy of Sciences (IAS) — which was founded in Bangalore in 1934 by his uncle C. V. Raman, who also won the Nobel prize for physics in 1930. He had previously been denied access to the archives of the Indian National Science Academy in New Delhi on the grounds that it "would revive old controversies".

In particular, documents showed that the government proposed the idea of a single academy in January 1947, and again, just days before independence, in August the same year. But the initiative failed, because each academy imposed "impractical" conditions.

Raman, who was president of the IAS until his death in 1970, was in favour of a single academy, but refused to agree to a transfer of IAS's property to the new body. This had been demanded by Shanti Swarup Bhatnagar, the president of the National Institute of Science — which was later renamed the Indian National Science Academy — as one of the conditions for his institute taking part in the merger.

The National Academy of Sciences in Allahabad also refused to hand over its property, and insisted that a "quota" of places on the executive body and fellowships of the new academy be reserved for members of the old academy. This proposal was rejected by Bhatnagar, and the scheme was abandoned.

Ramaseshan asserts that the proposal for a single national academy represented a lost opportunity for the Indian scientific community to have a single voice. As a result successive governments since independence have tended to take advice from individuals. This has amounted, he claims, to "showing patronage on a few and encouraging sycophancy".

K. S. Jayaraman

'Soft' privatization preferred for Natural Resources Institute

London. The British government is to privatize one of its major research institutes, the Natural Resources Institute (NRI) in Chatham, Kent. The institute, which specializes in research on natural resource management in developing countries, is at present run by the Overseas Development Agency (ODA).

The move follows last year's government 'efficiency scrutiny' of public research laboratories, which was partly aimed at identifying institutions suitable for privatization. The report has been strongly criticized by science bodies and trade unions, and more recently by a House of Lords select committee (see *Nature* 372, 718; 1994).

Although the review failed to come up with many candidates for privatization, it recommended that public research institutes should be 'routinely' transferred to universities, in what some consider a milder form of privatization than the sale of institutes to private companies.

Indeed, a consortium made up of the University of Greenwich, the University of Edinburgh and Wye College (part of the University of London) jointly submitted to the ODA last year a proposal to take over NRI.

But NRI members of the Institution of Professionals, Specialists and Managers (IPMS) — the main trade union representing the 530 staff at NRI — have protested against any transfer of NRI, and have written an open letter to the minister for overseas development, and have started a petition.

Rob Black, chairman of the ODA branch of the IPMS and a plant pathologist at NRI, says the short-term perspectives of universities are not compatible with the aims of NRI.

Others are concerned that the consortium has a head start on other potential bidders who will not have enough time to prepare proper bids. A shortlist of two or three bids will be drawn up shortly after the call for tenders is made public. Those shortlisted will then submit detailed formal tenders by April. **Maggie Verrall**

UK New Year honours

John Maddox, Editor of *Nature*, has been awarded a knighthood in the Queen's annual new year honours list for services to science. A knighthood also goes to Arnold Wolfendale, professor of astronomy at the University of Durham, and due to retire as Astronomer Royal this year.

César Milstein, deputy director of the Medical Research Council's Molecular Biology Laboratory in Cambridge is appointed a Companion of Honour.

1995: Science in search of 'strategies'

Almost a quarter of a century ago, a former British minister of education and science wrote a powerful — if controversial — article in *The Times*, where she asserted that “the party was over” for scientists. The minister was Mrs Margaret (now lady) Thatcher. Her epitaph was premature; science budgets continued to grow in most countries over the following two decades.

But in 1994, evidence from countries throughout the world suggests that science has indeed now lost its privileged place among political priorities, and that the scientific community is increasingly perceived as just another interest group seeking support from the public purse.

Funding for science has suffered because of the economic recession, while the end of the Cold War has reduced the political motivation of nations to seek supremacy in science and technology (see *Nature* 367, 5; 1994). But emerging trends also suggest that a deeper shift in political attitudes towards science is taking place.

The good news is that governments remain broadly supportive of science. Indeed, there seems to be a renewed awareness that a strong science base is essential to the welfare of a nation's economy, intellectual culture and quality of life.

What seems to be happening is that governments are now seeking new ways to mould the size and nature of their science base to respond better to such needs. The inherent risk of such exercises is that demands for greater accountability to the needs of both industry and the paymasters of science may result in science becoming orientated towards narrow and short-term goals.

On the following pages, *Nature* correspondents provide a brief overview of

such trends in the world's major scientific nations. In the United Kingdom, for example, the dialectic between intellectual advance and cost-effectiveness in science has intensified over the past decade, and is now playing itself out through the implementation of the recent white paper (policy document).

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Republican majorities mean changes for science in the United States.

In Japan, the contemporary pressures on science are different, and centre on the need to increase the accountability of researchers to their scientific colleagues, particularly outside Japan, rather than to government officials. But the goal is, as in the United Kingdom, to rationalize the way science is run in order to stimulate

intellectual creativity and maximize its contribution to economic growth.

France, Germany and the United States are also working towards this goal in their own ways. France and Germany seem to have rejected revolutionary changes, such as giving industry a markedly greater say in research, partly because of resistance from the research community. Instead, both countries have set about creating structures that will bring together scientists, industrialists and representatives of other sectors to try to seek a wide consensus on the best ways to proceed.

In the United States, the sweeping victory of fiscally conservative Republicans in November's mid-term congressional elections promises to accelerate changes in the organization and philosophy of science. This seems to indicate not simply the end of the honeymoon that science has enjoyed since the end of the Second World War — for example, through reductions in federal support for national laboratories — but also the start of a new social contract where science will be required to pay for its keep.

It is tempting to emphasize any convergence among various national trends over the past year. But each country's policy reflects its historical and cultural circumstances, and the changes being made in one country are not necessarily appropriate elsewhere.

Nonetheless, as we hope we have shown in the next few pages, there are enough common factors to suggest that some form of global consensus is now emerging as to how science might be organized to better meet the needs of society. The challenge is to negotiate such change without endangering scholarship and enlightened leadership. □

Science awaits the reality of Republican rhetoric

Washington. The 104th Congress of the United States convenes this week with Republican party majorities in both the House of Representatives and the Senate for the first time in 40 years.

The Republicans have promised large cuts in the \$508 billion spent annually at the discretion of Congress, which includes \$73 billion for science and technology. Examples of where these cuts might be were attached to the *Contract with America*, the ticket on which the party won the elections.

But prospects for the discretionary budget are much less certain than Republican rhetoric would suggest, for the resolve of the Republicans to make big spending cuts has still to be tested.

Following claims that they plan to pay

for tax cuts by adding to the budget deficit, the Republicans have promised to specify exactly where cuts will be made before they reduce taxes.

President Bill Clinton attempted to do likewise last month, promising large cuts at the Department of Energy and elsewhere, to pay for tax reductions for the middleclasses. But he stopped short of specifying which programmes would be closed, or in which states jobs would be lost. The Republican Congress may well resort to this familiar tactic, say observers, despite the party's claims to the contrary.

It is nonetheless clear that many science agencies will spend 1995 fighting as never before to defend their budgets, and in some cases to ensure their survival. Such impera-

tives risk preoccupying many agency heads this year rather than other more pressing tasks, such as revising scientific priorities or improving collaboration with other agencies and with the private sector.

If Republicans are intent on making large budget cuts, then all science agencies are potential targets. Increases in the budget of the National Institutes of Health (NIH), for example, have enjoyed strong bipartisan support over the past decade. But even the NIH's \$11-billion budget has now been identified by two influential right-wing think-tanks, the Heritage Foundation and the Cato Institute as a target for cuts of up to 15 per cent.

In fact, NIH is unlikely to face deep cuts. Instead, the impossibility of accommo-

► dating the growing number of suitably qualified biomedical researchers within a flat budget will remain its toughest problem.

The hit-list of cuts contained in the *Contract with America* would limit growth in the \$3-billion budget of the National Science Foundation (NSF) to 1 per cent less than inflation. Whatever happens to this list, the Republicans can be relied upon to prevent a repeat of last year's 14 per cent increase in NSF's budget, and to reverse the agency's recent drift away from basic research towards school education and applied research.

Fiscal conservatives are also likely to attack the amount of money that research universities spend on research overheads. What is not clear is how large such cuts will be, or how much mud will stick when Congress holds hearings to 'prove' that universities are wasting public money.

The Department of Defense is the only agency that Republicans want to enlarge. But defence hawks will continue to accuse it of wasting money on university and other research unrelated to battle-readiness. A proposal to bring a halt to such research — made recently by Senators John McCain (Republican, Arizona) and John Warner (Republican, Virginia) — may come to nothing, but the point they have sought to make will resonate in the new Congress.

The Department of Energy will probably be the favourite whipping boy of the new Congress. The department's political vulnerability is illustrated by the fact that Clinton chose it as the target for almost half of his own modest package of cuts. But his failure to specify exactly where the cuts would be made again serves as a reminder of the difficulties that Congress will face in the coming year in turning cost-cutting rhetoric into action.

One reason is that the US Constitution, which was drawn up by Virginian landowners, was designed to provide cautious, or even weak, government. Its intricate system of checks and balances — for example, the ability of senators from even small states to obstruct business in the upper house to ensure that their views are heard — rules out drastic changes such as those pursued by Mrs Margaret (now Lady) Thatcher after 1979 during her spell as UK prime minister.

So observers in Washington do not expect a revolution in the new year. Instead, they predict progressive tightening of the austere fiscal environment that has kept spending on science and technology more or less static since 1987, the year when four decades of growth came to an end.

Such a situation would leave the Republican Congress and the Democratic administration on familiar ground, blaming each other for the Congress's failure to deliver its promises to reduce taxes, enlarge the military, shrink government and balance the budget. Such 'gridlock as usual' will not impress the public, but it may be the best that science can hope for during 1995.

Colin MacLwain

Japan aiming to fund centres of 'excellence' in universities

Tokyo. Japan made a series of small but important moves towards reforming its public research system in 1994, to encourage the creation of centres of excellence in universities and government research institutes. The reform process, which will continue in 1995, is likely to have far-reaching effects on the development of science and technology in Japan.

Japan's 98 national universities come under the central control of the Ministry of Education, Science and Culture, which disperses most university funds in a bureau-

A new budget to promote centres of excellence was requested in August by the education ministry. This represents a major shift from the traditional practice of spreading money as evenly as possible among Japan's universities.

If the Diet (parliament) approves the budget request this spring, university-related institutes stand to gain an extra ¥3 billion (US\$30 million) annually, while a few small research groups would also each receive grants worth several hundred million yen (see *Nature* 371, 188; 1994).



A 'gang of four' university reformers? Yoshikawa (Tokyo University), Esaki (Tsukuba), Oda (formerly of ISAS) and Arima (RIKEN).

cratic, piecemeal fashion without regard for performance. A similar situation prevails in most national research institutes belonging to the education and other ministries. Under such a system, there is little incentive for institutions to excel. But moves for reform are underway.

The reform process began to take root last year with the publication of an external review of the Institute of Space and Astronautical Science — a world leader in X-ray and solar astronomy — carried out by leading scientists and experts from Japan and overseas. This review, initiated by Minoru Oda, a former director general of the institution, followed the first attempt to carry out such reviews in Japan, the previous year, by the physics department of Tokyo University and the Institute of Physical and Chemical Research (see *Nature* 365, 97; 1993).

In June, the Ministry of International Trade and Industry (MITI) announced plans to assess its 16 research institutes — but the reviews have yet to begin, because of resistance from the research institutes. At the end of last year, another external review, of the new graduate school of mathematical sciences at Tokyo University, was released.

The reformers responsible for establishing these reviews — such as Akito Arima, president of RIKEN and former president of Tokyo University — deny that they will directly influence government decisions on funding. But these generally favourable assessments of some of Japan's leading research institutions should nonetheless help them to attract extra funds.

The external reviews have all criticized the lack of foreign staff in Japan's leading research institutions. Robert Geller, from Tokyo University's faculty of science, is one of the few foreign scientists with tenure in Japan's universities. He has already spoken out about the problem (see *Nature* 372, 721; 1994), but considers that major improvements are unlikely because most Japanese universities recruit staff internally.

Until such problems are tackled, Japan's prospects for opening up its universities will remain poor. Indeed, the Ministry of Education, Science and Culture seems to be moving backwards. It is recently reported to have issued a directive encouraging universities to terminate the contracts of senior foreign faculty (over 45) to save money on the higher salaries they command.

In May, Tsukuba University created the Tsukuba Advanced Research Alliance (TARA) to encourage cooperation between its faculty and researchers in industry and government research institutes in Tsukuba science city (*Nature* 369, 268; 1993). TARA faces many bureaucratic obstacles, but Leo Esaki — president of Tsukuba University and the 1973 Nobel prizewinner in physics — is determined that it should succeed; he is supported by senior officials in the education ministry.

Similarly, Hiroyuki Yoshikawa, president of Tokyo University, has taken a step towards providing the university, with a greater say in its own financial affairs by setting up a 'presidential fund' (see *Nature* 372, 394; 1994).

David Swinbanks

Germany courts consensus not controversy

Munich. The future of basic research is probably safer in Germany than in any other European country. Pressure for greater emphasis on applied research is counterbalanced by such pillars of the basic research community as the Max Planck Society (which runs medium-sized research institutes), the large national research centres and the Deutsche Forschungsgemeinschaft (DFG, which funds university research in universities).

Indeed, these politically independent organizations successfully resisted the biggest post-war attack on basic research in Germany, which took place two years ago. This attack was prompted by the high costs of both re-equipping laboratories in the five new *Länder* and East Berlin, and assimilating them into the West German system, following reunification in 1990.

These costs prompted a shift in the political debate on research at the end of 1992, away from issues directly concerning the East towards Germany's research system itself.

In particular, the government began listening to demands that public research should be more orientated to meeting industry's needs. Hans Zacher, president of the Max Planck Society, recalls 1993 as "a very bad experience". "It was as if Birnam wood had started to march on research, and in the wood were the combined troops of government and industry." Basic researchers had to justify their work at endless meetings convened by industrialists, he says, while the government considered devoting some national research centres to industry's service.

The political mood has shifted, however, and is now more reflective than aggressive. While the government recognizes that the transfer of technology from the country's strong science base to industry needs to be improved, the research ministry has toned down the more extreme demands of industry, summarized in the so-called Weule report (see *Nature* 372, 4; 1994).

The ministry of research has also rejected the idea that industry should have more direct control over the direction of state-funded research. In typical German fashion, the ministry has instead set about searching for the best means for research organizations, industry and government to reach a consensus on how to proceed.

For example, Paul Krüger, the former research minister, set up a 'strategy circle' of 14 scientists and industrialists to advise him on how technology transfer might be

improved. He also set up a series of more focused 'dialogue circles' to identify areas of research that could directly benefit industry.

The ministry also convened regular meetings with the 'holy alliance', which includes the presidents of the Max Planck Society, the German Rectors conference, the AGF (the umbrella body of the national research centres), the Fraunhofer Society (which runs applied research institutes), the DFG and the Wissenschaftsrat—the German research council.

Last summer, the strategy circle recommended improving the legal environment for research—for example, by reducing the bureaucracy faced by biotechnology researchers—encouraging private support for research through tax incentives and developing special research programmes in state-funded laboratories in collaboration with industry. Discussions on strategy should also be continued at the ministry level, it added.

The German cabinet also approved plans last year to set up a high-level technology council made up of scientists, industrialists and representatives of the trade unions and ministries, and chaired by Chancellor Helmut Kohl. But this failed to take off at the time, because Kohl wanted to delay creating the council until after last October's general elections.

But the renewed appreciation of the importance of research and education to Germany's economy was emphasized again after the elections when the ministries of

research and education were combined into a powerful "ministry for the future". Responsibility for the technology council was also transferred from the cabinet office to the new ministry; Jürgen Rüttgers, the new science minister, intends to hold its first meeting next month.

Rüttgers is also keen to establish what he terms a "German national academy of sciences" to supply independent advice on scientific issues. The plans have been controversial in the academic community, however, because the academy would include not only scientists but also representatives from industry, all political parties and trade unions.

In particular, the seven small local academies dotted around Germany feel particularly slighted—a real academy does not have this sort of composition, says Hubert Markl of the Berlin-Brandenburg Academy. The academies have written jointly to Rüttgers offering to provide the advisory service that he seeks.

The 'strategy circle' is also to be replaced by a holy alliance, which will be extended to include the presidents of the Humboldt Foundation and the Deutscher Akademischer Austausch Dienst. These organizations offer German fellowships and studentships to foreigners.

How all these various advisory groups will work and interact with one another has not yet been outlined. But Rüttgers wants them to be up and running within the next few months.

Alison Abbott



Rüttgers: wants advisory groups.

Science needs new role in new South Africa

Cape Town. South Africa's research community will start the new year amid uncertainty about how much the government intends to spend on science and which areas will be supported.

Ben Ngubane, the minister for Arts, Culture, Science and Technology, will face a difficult task in persuading the cabinet to increase the R886-million (US\$252-million) science budget in the face of other priorities.

Morale is low among scientists, and in particular basic researchers. The Foundation for Research Development (FRD), the government's agency for science and engineering research, will cease funding grant-holders at the end of the year, and invite them to apply for new programmes due for launch soon after.

These will include an 'open programme' for basic research, according to FRD, but the other programmes will be orientated towards industrial and social objectives: "competitive industry, improved quality of life, effective science, engineering and technology education and

awareness," and "sustainable environment".

Olive Shisana, the special adviser to Nkosazana Dhlamini Zumathe, the health minister, also told deans of medical faculties, at a meeting last month, that medical research priorities should be directed towards improving access to health care. Most medical research is also funded from academic hospital budgets, but these are likely to be reduced in the next financial year, and the savings diverted towards primary and secondary health care.

There is also concern that the R15 million that the Medical Research Council provides for university-based fundamental research may be under threat. "Even to contemplate eroding this minuscule amount is irresponsible," says Bob Millar, assistant dean (research) at the University of Cape Town Medical School. This money is particularly important in the training of medical researchers, he says, because it funds research in areas other than national health priorities.

Michael Cherry

British white paper brings 'cultural change'

London. Research is being described in new terms at the Engineering and Physical Sciences Research Council (EPSRC), the largest of Britain's six research funding agencies. The council increasingly refers to university-based research groups as its "suppliers", and industry as its "customer base".

Alan Rudge — the chairman of the council, and former director of research at British Telecom — goes further, describing the council as similar to a board of directors, with responsibility for defining strategy and providing "sensible answers" to problems that the council faces.

The council's new thinking is representative of one of the most striking changes catalysed by the government's 1993 white paper (policy document), *Realizing our Potential*. (The council itself was created following a recommendation in the paper to break up the Science and Engineering Research Council (SERC).)

Government officials claim that this shift in self-perception is an important part of a cultural revolution, which they have sought to engineer through the changes outlined in the white paper. These emphasize using the science base to create wealth and — although in practice to a lesser extent — to improve quality of life.

The white paper has had less impact in other areas, however. The threat that the Medical Research Council's institutes might be sold or privatized, for example, has not materialized. The agency — which, observers point out, already works closely with industry and the health services to develop applications of its research — has been able to continue operating much as it did before.

The Particle Physics and Astronomy Research Council (PPARC) — which was also created from the break-up of SERC — also continues to fund mainly basic research. It is under increasing political pressure, how-

ever, to get better value from its subscriptions to international facilities (see *Nature* 372, 712; 1994).

Moreover, the government has dropped some of the most radical — and widely criticized — reforms proposed in the white paper. For example, the new one-year master's degree — billed as a cost-effective initiation in research — will not be compulsory for students wishing to apply for PhD programmes funded by the research councils, as previously proposed (see *Nature* 372, 7; 1994).

Similarly, the government is likely to shelve proposals to privatize many of its public research laboratories. A civil-service review intended to prepare for their sale has been criticized by almost all the bodies interested in the health of British science (see *Nature* 372, 718; 1994).

The outcome of the ambitious 'technology foresight' exercise, which the white paper proposed as the major channel of 'user input' into government science policy, is more difficult to predict. Early fears that it would focus on narrow industrial goals have subsided, but many remain concerned that pressure from the Treasury for results from public investment in science could lead to emphasis on short-term thinking.

The view that the most important impact of the white paper has been to prompt an evolution in attitudes in the research community is also shared by Tom Blundell, chief executive of the Biotechnology and Biological Sciences Research Council

(BBSRC) — which combines the activities of the former Agricultural and Food Research Council with SERC's biological work — and professor of crystallography at Birkbeck College in London.

Blundell says that there has not been a shift from basic research to either strategic or applied research, as some had feared. The white paper, he argues, has helped to focus attention on the importance of basic science to long-term questions of wealth — and health — creation.

Both the EPSRC and the BBSRC are faced, to a far greater extent than the other research councils, with bridging the development gap between technologically conservative industries — for example, the agrofood industry — and the innovative potential of discoveries in basic science.

Both councils are also struggling to introduce new procedures for allocating research funds. Their aim is to adapt the best of former practices for selecting and assessing research grants to their new pressures and priorities. The consequent changes in the role and structure of the peer-review process will be a contentious issue over the coming months.

The most urgent task for both councils is to reconcile the government's demand to mould research more closely to national needs and reduce spending. Rudge says it is his council's responsibility to carry out the government's instructions as "efficiently and effectively as possible". Rudge considers final staff numbers as an "output of the way of doing things right".

A different view is taken by the Institution of Professionals, Managers and Specialists, the labour union that represents many government-funded scientists and science administrators. The union opposes cuts and argues that a sufficient critical mass of science expertise is needed to preserve a strong science base.

Events this year will help to clarify the impact of the white paper — in particular, the new research council structures will be put in place and the government will publish the first conclusions of the technology foresight exercise.

But conflict can also be expected as a result, for example, of the large staff cuts expected at the research councils and demands for greater cost-effectiveness in international programmes, such as those financed through the European Union's Framework programme, or the European Space Agency.

David Dickson

Correction

Lyle Bivens, director of the US Office of Research Integrity (ORI), was inadvertently referred to as "she" in a recent news story (*Nature* 372, 391; 1994) about the ORI's report on a paper by Thereza Imanishi-Kari.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

**Blundell: white paper
has emphasized basic
research.**

Indian science turns to industry

New Delhi. The past year in India has seen government laboratories being forced to turn increasingly to industry for support, because of reduced public funding. This, in turn, has put pressure on basic research which has given way to commercially exploitable applied research.

Another factor that has encouraged Indian scientists to commercialize their results is a new law allowing product patents in health, food and agriculture, following the signing last year of the General Agreement on Trade and Tariffs (GATT). The liberalization of economic policies has also helped to encourage commercialization of research results.

As evidence of this, India exported — or licensed or patented abroad — several technological products in 1994. "India is no longer just an importer of technology," says S. K. Joshi, secretary to the Department of Scientific and Industrial Research. "We have become a global player in some areas."

The research and development culture in industry is also changing because of the opening up of India to foreign technology and capital. In particular, drug companies are increasing investment; the New Delhi-based company Ranbaxy Pharmaceuticals Ltd, for example, has set up a US\$5-million centre for drug research. **K. S. Jayaraman**

France: too much analysis leads to paralysis?

Paris. French science faces a year of upheaval and confrontation in 1995, with the government seemingly intent on tackling the taboo of reforming both the university system and the country's largest research organization, the Centre National de la Recherche Scientifique (CNRS).

The moves are part of the government's plans for a new national strategy for research, based on the conclusions of a vast 'national consultation' on research that it organized last year (see *Nature* 369, 599; 1994). This urged France to "rediscover a strategic vision", and review methods for running research in the light of the economic and political changes of the past decade.

One argument is that the rapid growth in the scope of research possibilities means that choices must increasingly be made. This trend has been sharply reinforced by the freeze or reduction in research budgets imposed by the economic recession. Pressure is also mounting on scientists to justify spending and to provide better value for money.

Moreover, the general political trend towards greater liberalization and deregulation is leading France, like other countries, to review its policies for providing public aid to industry.

The government also claims that the public wants research to be more orientated towards social objectives, such as boosting economic competitiveness, reducing unemployment and improving the environment. The national 'strategy', it says, will allow research funds to be better matched to public policy needs, for example in health, transport, telecommunications, agriculture and energy. But the French debate is less about whether a strategy is needed than about redefining the respective roles of the government and research bodies.

Indeed, while the United Kingdom and other countries are for the first time seeking ways of planning research in the long-term, the French have been doing so for years.

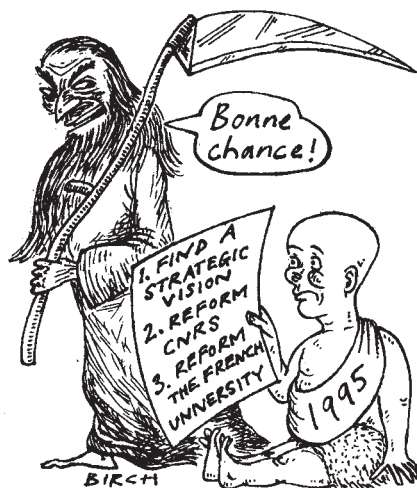
Long-term policies in many sectors, including research, are regularly proposed, for example, by the powerful Commissariat général du Plan, a commission made up of industrialists, trade unions and representatives of various public organizations. Such forward planning has led, for example, to the creation of the *grands programmes* in space, nuclear science and aeronautics, and has provided successes such as the Ariane launchers and the Airbus airliners.

But such government *dirigisme* is now coming increasingly under fire. The national consultation concluded, for example, that the research system is "too rigid" and that the government intervenes in research too directly.

François Fillon, the French science min-

ister, argues that the government should have a more hands-off approach and concentrate on "initiating, coordinating and stimulating".

At present, the French research system is bureaucratic. Most researchers are civil servants with life-time tenure, working for one of the research organizations, such as the



CNRS — which employs 12,000 researchers and 15,000 support staff — or INSERM (see *Nature* 363, 751; 1994).

But the government has ruled out adopting the 'supplier-customer' model of research administration, which is favoured, for example, in the United Kingdom (see page 8). Observers argue that the model is flawed, because, although it assumes that the market can define research directions, it can do so only for short-term objectives. It cannot provide long-term strategy.

Indeed, some argue that 'foresight' exercises, which are currently fashionable in the United Kingdom and elsewhere, are a tacit acknowledgement that research policy cannot be driven by the market, in that their implicit aim is artificially to create demand and supply.

The emerging consensus in France is that the government's role should be to impose strict priorities for areas such as large equipment, technology transfer and applied and technological research, while leaving research organizations considerable freedom to choose which areas of basic research to support. Fillon says that to make long-term plans for basic research or innovation would be "illusory".

The structure proposed to articulate the government's strategy is a national committee for strategic orientation within the ministry of higher education and research. The committee — which will be made up of scientists, industrialists and representatives from other sectors, appointed by the prime minister — will also submit its recommendations to the National Assembly before the

annual budget vote.

Fillon also plans to launch a system of five-year contracts with the research organizations, fixing objectives and the means to achieve them (see *Nature* 365, 196; 1993). Counterintuitively, this will probably result in less central control, not more. Contracts, says one observer, are a means by which the research organizations "negotiate" medium-term objectives with the state in exchange for "liberty and autonomy".

One likely consequence is the transformation of the CNRS from a research organization employing thousands of researchers into a research council concerned primarily with distributing grants. At present, the agency allocates most of its research funds directly to its laboratories, but one proposal would shift much funding towards new 'programmes', where grants would be awarded on the basis of competitive proposals (see *Nature* 371, 639; 1994).

One problem is that the CNRS has little money to fund its proposed programmes because it spends more than three-quarters of its budget on salaries and overheads. To increase its margin for manoeuvre, the CNRS may reduce the number of its laboratories — possibly from 1,300 laboratories to as few as 400, with control of many passing to the universities. Such a move would be highly controversial.

In particular, although France has many first-class universities, the system is already hard pressed to recruit the staff needed to keep pace with burgeoning student numbers, and might not be able to afford to maintain new laboratories adequately.

Many argue that the universities need to be reformed if the CNRS is to be restructured successfully. That would be politically explosive. Successive left-wing governments have attempted to reform the university system, while keeping its nationalized status — the most recent with some success. But the right has regarded privatization as a remedy. Both have been impeded by the resistance to change of powerful lobbies of students and *professeurs* (see *Nature* 364, 5; 1993).

But the need to "reinvent" the French university is one of the main recommendations of a recent report, commissioned by Edouard Balladur, the prime minister, on the challenges facing France over the next decade. The report claims, in particular, that the universities lack leadership and policies, and that their systems of evaluation and management are "mediocre".

The report recommends that universities should be allowed to hire administrators from outside. It also asserts that "it is time to lift the taboo" on charging fees. "Free education is not providing equality", it argues, but "impoverishing" the universities.

Declan Butler

Popper's achievement

SIR — I should like to add to Hermann Bondi's obituary of Sir Karl Popper (*Nature* 371, 478; 1994). Popper cleared away an error that had entrapped ideas about human rationality since the Greeks.

The rationality of science poses a problem: scientists judge some theories to be better than others. In doing this they reason with arguments. But how do they give that judgement rational authority? From Plato onwards, one approach has dominated: philosophers took the authority of rationality to be in some way transmitted from a source. Popper broke with this tradition. Rationality, he argued, got its authority from our goals — in the case of science — the search for truth. Likewise for the rationality of politics: no source exists for the right rule of society, but we can aim for better ones.

The novelty of his innovation stands out against the history of attempts to fathom the nature of human rationality. This has otherwise been one of many failed attempts to find its source. Plato suggested Forms and Ideas. Others have proposed God, Being, innate or *a priori* ideas, empirical data, monads, elementary propositions and various types of intuitions and foundations. Sometimes the notion that rationality is linked with truth has been dropped while its origin in a source has been kept: ideological superstructures, social contingencies and epistemes. Even sceptics did not deny that the authority of rationality came from a source, only that it could be transmitted to our beliefs.

This pursuit arises from a basic error: the assumption that because we use sources to back our arguments, the ultimate authority made by arguing — rationality — must likewise come from one. But counter-examples exist that dissociate them. Consider air safety: aircraft designs have an authority — airworthiness — that derives from argument and sources — research findings and crash reports. But the authority they make comes from a goal — the desire to fly safe aircraft. No argument can show that an aircraft is safe (we cannot inductively infer that an accident-free aircraft will be so in the future). What argument can do is to let us (or rather civil aviation bodies on our behalf) winnow out unsafe aircraft designs. The more vigorous and exhaustive that search, the more authoritative such judgements are about airworthiness.

The source approach to rationality is also rooted in human experience. As people, we often seek sources of authority to guide what we should do and believe. Children when uncertain look to their parents (social referencing). Uncertain adults, as Stanley Milgram found in his infamous obedience experiments, turn

not to their own judgement but to apparent figures of authority. People in religion seek the origins of morality, existence and meaning from supernatural sources. Philosophers often sought in metaphysics a realm of certainty to which they could turn to satisfy such needs for rational beliefs.

That was until Popper. He was the first thinker to explore whether rationality could come from our goals. Whether his suggestions of how we did this — criticism, conjectures and refutations — were the right ones, he asked questions that philosophers should have been asking centuries ago. By venturing to raise and explore them in the twentieth century, he, more than anyone else since the Greeks, changed how we understand the function of reasoning in science and society.

John R. Skoyles
8 Denning Road,
London NW3 1SU, UK

Origins of steady-state theory

SIR — Whatever may be thought of it now, the steady-state theory of the expanding Universe was an important event in the evolution of cosmology. As there are somewhat contradictory statements about its origin in the literature, we thought it advisable to put together our joint recollections of the history of those days while all three of us are still available.

F.H. awakened a deep interest in astronomy in H.B. and T.G. during our period together during the war, and frequent discussions on questions of astronomy were a central feature of our lives after all three of us returned to Cambridge in 1945–46.

A pressing problem of those days was the discrepancy between Hubble's estimate of his constant and other timescale considerations, as well as questions of the origin of galaxies and of heavy elements. It was T.G. who first proposed the continuous creation of matter at a low but sufficient rate, making it possible to have a universe whose large-scale features remain constant in time. Among the advantages this offered was the removal of the discrepancy between local and cosmological timescales. All three of us debated this idea exhaustively and were then greatly taken with it. But was it publishable merely as an idea?

F.H. then (early 1948) formulated the first version of his C-field theory and offered T.G. co-authorship, which T.G. declined, not wanting to have the first publication tied to so particular a formulation. F.H. then wrote it up as a paper for publication with an acknowledgement to

T.G. Meanwhile, H.B. and T.G. refined their ideas and devised the perfect cosmological principle as well as studying the refutability of their theory in the sense of Popper. Close contact was maintained with F.H. and many discussions were held about this and other topics in astronomy, with F.H. expressing some doubt about H.B. and T.G. using so philosophical a basis.

In spring 1948, H.B. spotted an observational agreement that made publication appropriate. H.B. and T.G. immediately wrote up their phenomenological paper, with much emphasis on the thermodynamics of such a steady-state Universe. This paper (*Mon. Not. R. astr. Soc.* 108, 252; 1948) appeared just in front of F.H.'s paper (*Mon. Not. R. astr. Soc.* 108, 361; 1948) in the autumn. This latter paper had been turned down in the spring of 1948 by a well-known physics journal on the curious ground that it was experiencing a shortage of paper.

Hermann Bondi
Churchill College,
Cambridge CB3 0DS, UK
Thomas Gold
Center for Radiophysics,
and Space Research,
Cornell University,
Ithaca, New York 14850, USA

Fred Hoyle
102 Admiral's Walk, West Cliff Road,
Bournemouth, Dorset BH2 5HF, UK

Research integrity

SIR — An article headed "Study proposed on integrity of published data" (*Nature* 371, 733; 1994) states that Congress's decision to require the National Institutes of Health (NIH) to set up a commission on research integrity was prompted by frustration with the way in which allegations of scientific misconduct have been treated and concern that the NIH's Office of Research Integrity (ORI), which investigates allegations of misconduct, has not been working as well as Congress had hoped.

Although the Congress called for the commission through language in the NIH Revitalization Act, the commission itself was set up by the Department of Health and Human Services (DHHS), with members selected by the DHHS Secretary, Donna Shalala. Moreover, the ORI is not and never has been a part of the NIH. It was established by the Public Health Service (PHS) on 3 June 1992, when NIH's existing Office of Scientific Integrity was combined with an existing PHS-level office, the Office of Scientific Integrity Review.

Anne Thomas
DHHS Public Health Service,
National Institutes of Health,
Bethesda, Maryland 20892, USA

Eighteen ninety five and all that

J. L. Heilbron and W. F. Bynum

Anniversarial commemorations this year include the discovery of X-rays, the calculation of the position of the undiscovered planet Neptune, the emergence of syphilis and the mandatory serving of lime juice to British sailors.

DURING the Christmas season of 1895, Wilhelm Konrad von Röntgen made the hand he had taken in marriage the most famous hand in the world. A photograph of its inside taken with his "new sort of ray" sent physicists back to their laboratories with (we suppose) their own wives, to try the experiment; which was readily done, as most of them had the necessary equipment, with which they had been generating weak X-rays unknowingly for years. Röntgen's ray quickly became an indispensable finding aid in medicine. On physics too it had an immediate and profound impact. As a laboratory tool it led directly to the definitive detection of the electron; as an indicator that nature still had secrets undreamed of by natural philosophers, it helped to burn away whatever lethargy and complacency there may have been in *fin-de-siècle* science.

It took seven weeks, from 8 November to 28 December, for Röntgen to go from first inkling to publication of his discovery. Midway through this period, on 25 November, Alfred Nobel signed the will endowing his annual prizes for "those persons who during the previous year have rendered the greatest service to mankind". This prescription fitted perfectly the results Röntgen achieved as it was being written. He duly received the first Nobel prize in physics.

Röntgen qualifies not only for a centennial (1¢) for his rays, but also for a sesquicentennial (1.5¢) for his birthday. That he made his only important discovery at the age of 50 should give heart to middle-aged physicists who have not yet found anything. He is our anniversarian of the year.

As runner-up we choose an invention of 1745 (2.5¢). Like X-rays, the Leyden jar came forth through manipulation of standard electrical apparatus of its day and so was easily and immediately repeated. Again like X-rays, its properties challenged existing physics — in its case, all previous electrical theories — and it had a medical application. Its shock, taken in moderation, cured paralysis, flatulence and arrogance. More below, under 1745.

There are many other events in science worthy of, and entitled to, commemoration this year. We list them backwards, beginning with centennials, sesquicentennials and so on, down to 1495. We then return to our century, to semicentennials and to our popular 50 (± 25)-year observations.

1895 (1.0 centenary)

In preparing to enter what for him was a new research field, Röntgen repeated all the chief experiments on the discharge of electricity through gases. One of these experiments, developed by Philipp Lenard, employed a fluorescent screen to detect rays of uncertain character that

Engineering in St Petersburg, to implement repressive tsarist measures against the students.

A more significant milestone for women in science than Mrs Röntgen's manual contribution was the awarding to Grace Chisholm Young of the first PhD earned by a woman at a German university. The university was Göttingen and the subject mathematics. Two years later, prominent German academicians were asked about the desirability of admitting women to study for higher degrees. Max Planck's answer represented the moderate-liberal point of view. Women should be allowed to attend lectures "as far as it is compatible with academic order" and without encouraging them to step outside the sphere that "Nature itself has prescribed for [them]". Another stage on the hard road to women's access to higher education is noticed under 1920.

While near the subject of mathematics, we observe that, in 1895, Georg Cantor (born 1845, 1.5¢) published his magnum opus on set theory, *Beiträge zur Begründung der transfiniten Mengenlehre*; that Jules Henri Poincaré's *Analysis situs* founded modern topology; that Giuseppe Peano explained in only five volumes, in his *Formulario matematico*, how to re-express arithmetic and geometry as formal logic; and that Arthur Cayley, whose *Theory of Linear Transformations* had elevated the study of algebraic invariants when it appeared in 1845 (1.5¢), died.

Death also claimed the century's most famous scientific icon, Louis Pasteur. He was interred in the Paris institute that bears his name and was further perpetuated a century ago by the founding of similar eponymous establishments in Lille and Indo-China. But for all that, we prefer, for the institutional anniversarial prize, the foundation of the New York Public Library, not merely because it is a major repository of knowledge but also because its first director, John Shaw Billings, was the medical man who had conceived and edited, from 1879, that instrumental tool of biomedical research, *Index Medicus*.

We are not done with obituaries. In 1895, the grim reaper also reaped Darwin's bull-dog, Thomas Henry Huxley; Johann Joseph Loschmidt, a founder of the old Vienna school of particle physics, whose measurement of Avogadro's number so impressed his countrymen that they call it Loschmidt's number; Felix Hoppe-Seyler (born 1825, 1.7¢), pioneer German

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Inner visions — a caricature of Röntgen, discoverer of X-rays, published in *Lustigen Blättern* around 1900.

Lenard had reported finding outside the discharge tube and within a few centimetres of it. Röntgen left the screen on a chair as he warmed up his apparatus. It glowed far beyond the range specified for Lenard rays. Chance favours the prepared mind. Röntgen locked himself in his laboratory to study his "neue Art von Strahlen". Some weeks later he invited his wife in for a photograph.

At the other end of the scale, electromagnetically speaking, Guglielmo Marconi and, independently of him, Aleksandr Stepanovich Popov, invented the antenna as a detector of radio waves. During the next few years Marconi developed his circuit into a practical device, which by 1901 could detect a weak signal across the Atlantic. Like Röntgen, Marconi received the Nobel prize in physics for his service to mankind. Popov got nothing, having died in 1906, three years before the award to Marconi. Some think that he died from high tension generated by his refusal, when director of the Institute of Electrical

biochemist, best known for his researches on haemoglobin, which he discovered in 1864; and Julius Lothar Meyer, a physician turned chemist, who almost discovered haemoglobin himself and went on almost to invent the periodic system of the elements. On the other side of the ledger, Lancelot Hogben, who was to bring mathematics to the millions, and Robert Buckminster Fuller, the colourful American inventor of the geodesic dome and eponym of buckminsterfullerene, were born.

Other inventors alive in 1895 included the Lumière brothers, Auguste and Louis, who that year created the first movie (of workers in their factory) with their cinematograph; C. T. R. Wilson, inspired to reproduce the weather of the Lake District in the Cavendish Laboratory at Cambridge, built the first, primitive cloud chamber; and, Konstantin Eduardovich Tsiolovsky, a self-taught aeronautical engineer before the advent of aeroplanes, began to write about interplanetary rocket travel.

1845 (1.5 centenary)

Planet Uranus's habit of straying from its calculated orbit was an embarrassment to responsible astronomers. At least three plausible explanations offered themselves: there was an error in the calculations; Newton's law of gravity did not hold so far from the Sun; or an unknown something kept the planet from its path. In 1845 powerful computers on both sides of the English Channel, John Couch Adams and Urbain Leverrier, sure of their calculations and almost as sure of Newton's law, decided on a perturbing planet and gave its position. Neptune came to light the following year. The obvious moral of this tale was lost around 1915 (0.8¢) when Einstein explained discrepancies in Mercury's orbit, which some astronomers had attributed to an unobserved planet they called Vulcan, on the premise that Newton's law does not hold so close to the Sun.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Perturbing planet — false-colour image of Neptune, recorded by Voyager 2 in August 1989.

While Adams and Leverrier calculated, Karl Ludwig Hencke was continuing his lengthy search for asteroids. In 1845 he found one, which he called Astraea after the Greek goddess of justice, thinking, perhaps, that it was only just that he should have found it. The first four asteroids were discovered just after 1800; then nothing, for more than 40 years. Astraea started a fad, which resulted in the detection of well over 300 asteroids by 1895. The story is continued under 1920.

Astraea and Neptune do not exhaust the possibilities for sesquicentennials of interest to astronomers. In 1845 William Parsons, Earl of Rosse, finished building his four-ton telescope and started a search of the heavens that brought the first sightings of spiral nebulae. And Edgar Allan Poe continued his flirtation with darkness with a solution to Olbers' paradox, "Why is the sky dark at night?" (If the Universe is infinite and isotropically populated with stars, their light should fill the sky.) Poe's answer: the Universe has a finite age.

Further to the domestic touch of our anniversary of the year, Christian Friedrich Schönbein, a *naturphilosophisch* (that is, romantic) chemist and discoverer of ozone, tidily wiped up a laboratory spill with his wife's cotton apron. It went up in a puff of smoke. Guncotton subsequently had some play as an explosive, particularly in England, where Sir Frederic Abel, who held a patent on it, used his position as government adviser to have Nobel's nitroglycerine banned in Britain. To collect enough of his explosive to demonstrate its relative safety to the authorities, Nobel imported it in crates marked "light wine from the Rhine country".

Another chemical breakthrough worth celebrating was Adolph Wilhelm Kolbe's synthesis of acetic acid entirely from nonorganic ingredients, which drove a second nail (urea was the first) into the coffin of the vitalistic distinction between ordinary substances and those produced by living beings. Some rhapsodical souls anticipated synthesizing life. For further encouragement, see under 1970 below.

The ever-productive Michael Faraday had a banner year in 1845. It was not enough for him to discover para- and diamagnetism and to interpret them through his field theory. The same theory also helped him to persevere in searching for a relationship between light and magnetism, which he found in the rotation of the plane of polarization of polarized light passed

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Keeping scurvy at bay — the serving of lime juice to British sailors, as shown in *The Graphic* (1875).

through a magnetic field.

The microscope demonstrated its worth in medical diagnosis a sesquicentennial ago. In Edinburgh, John Hughes Bennett (d.1875, 1.2¢) described a patient whose blood had far too many white cells. He called the condition leucocythaemia. In Berlin, Rudolf Virchow wrote up six weeks later his own patient with a similar dyscrasia of the blood. He called the condition leukaemia. Virchow was the better microscopist, Bennett the better linguist.

Scientific American, the Royal College of Chemistry and the digs at Nineveh were born in 1845. So was Zygmunt Florenty von Wroblewski, pioneer of low-temperature research, who, with his colleague K. S. Olsewski, first liquefied oxygen, nitrogen, hydrogen and carbon monoxide, and who might have discovered superconductivity had he not overturned his kerosene lamp and, in a cruel irony, perished by fire.

1795 (2.0 centenary)

We have two food items to offer. For one, in 1795 the British Admiralty made mandatory the serving of lime juice on its ships at sea. The treatment kept the British tar, or limey, relatively free from scurvy until undone by technological progress. A prime mover in this progress was a Frenchman, Nicolas-François Appert, who, picking up a challenge thrown down by Napoleon, also in 1795, devised a way to preserve large quantities of fruits, juices and vegetables in jars. (He also invented the bouillon cube.) The same process when applied to limes destroyed most of their vitamin C. Much to the puzzlement of the Admiralty, scurvy returned to British ships. It arrived at the obvious erroneous conclusion that lime juice is not an antiscorbutic. For more about vitamin C, see under 1970.

James Hutton's *Theory of the Earth, with Proofs and Illustrations* dates from 1795. As the bible of uniformitarian geology and an inspiration for Charles Lyell's

Principles of Geology (1830), Hutton's *Theory* may be the pick of this year's bicentennial subjects. The French Republic re-established in 1795 the Académie des Sciences it had liquidated the previous year as the first class of a new global collection of savants, the Institut de France. No doubt some will prefer its bicentennial to Hutton's.

Among other comings and goings we note the birth of Christian Gottfried Ehrenberg, scientific traveller and microscopist, and the death of Josiah Wedgwood, experimental ceramicist and inventor of the porcelain from which, we think, many Enlightenment savants took in their material parts.

1745 (2.5 centenary)

In 1745 Pieter van Musschenbroek described to his correspondent in Paris his sensations on experiencing the shock from a Leyden jar. "I thought I was done for. . . I would not take another for the entire Kingdom of France." Heeding this warning, and contributing to our domestic theme, Johann Heinrich Winkler, a professor in Leipzig, short-circuited his jar with his wife.

Musschenbroek received a shock to his mind as well as to his body. The jar jarred most powerfully under circumstances in which, according to received theories, it should have been weakest. A retired printer from Philadelphia, Benjamin Franklin, rushed into the vacuum. His explanation made action over macroscopic distances subject to experiment and opened the way to the measurements of C. A. Coulomb in 1785 (2.1¢). The shock of the jar, which made paralysed limbs tremble, gave hope that after a course of such treatments the owner of the limb would be able to move it purposefully. Few if any uncontested cures were achieved, even by London surgeons who effected their guild separation from the barbers in 1745.

Like most French-Swiss savants of the

eighteenth century, Charles Bonnet was related to all the great families of Geneva. His most important discovery related to a family of 95, produced without a male. They were the offspring of his favourite aphid. His description of her parthenogenic feat appears in his *Traité d'insectologie* of 1745. The same year he published several monographs on the regeneration of rainwater worms, one of which, cut into 26 parts, became 26 worms. Their easy and natural multiplication offers an arresting contrast to Jacques de Vaucanson's artificial flute player, which employed the most intricate mechanisms to mimic the motions and breathing of a human performer. Vaucason qualifies for mention here not because of his flautist, but because of his self-acting silk loom of 1745.

1695 (3.0 centenary)

This year 300 years ago died one of the most accomplished natural philosophers of the Scientific Revolution, Christiaan Huygens. Discoverer of the rings of Saturn through a telescope of his own design, inventor of a truly isochronous pendulum and clock and, at the end of his life, historian of the people on the Moon, Huygens was called to head the Académie des Sciences in Paris from its beginning despite the considerable disability of being a Dutchman. Since Huygens is to the Netherlands what Galileo is to Italy and Newton to England, we anticipate widespread and illuminating commemorations of his life and work this year.

1645 (3.5 centenary)

We understand that geometry owed its invention to the Egyptian tax collector, who assessed liability on the basis of the property that remained after the annual flood of the Nile. The French tax collector, in the person of Blaise Pascal's father, inspired the first successful mechanical calculating machine. Distressed at the sight of his father labouring over his accounts, Pascal devoted five years, on and off, to designing a mechanical means for performing addition and subtraction. He succeeded in 1645 and set up a small business. Seven of his machines survive. That reminds us that a mass-produced mechanical calculator called the Millionaire went on sale in 1895 (1.0¢).

Also in 1645, Daniel Whistler wrote up his Leyden MD thesis on rickets. Although it contains the first description of the disease, so low was the subsequent reputation of this embezzler and sometime president of the Royal College of Physicians of Lon-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Huygens — multi-talented astronomer and physicist who died 300 years ago.

don that Whistler has been accused of plagiarizing from a book published five years after his own. Since we now know that rickets is caused by a lack of sunlight, the title of Whistler's thesis, *De morbo puerili Anglorum*, was prescient. For more on the naming of diseases, see under 1495.

1595 (4.0 centenary)

The field of quadricentennials belongs to Gerardus Mercator's posthumously published *Atlas sive cosmographicum*, a compilation of regional and world maps that gave its name to the genre. It may stand for the brilliant cartographical work carried out in the Low Countries during the sixteenth century, which literally gave the world its present shape — apart from a river around the North Pole, a huge ice-bound Antarctica and an island named California. The Mercator projection, which misrepresents the globe at high altitudes in order to make loxodromes into straight lines, dates from 1569.

1545 (4.5 centenary)

Girolamo Cardano was a mathematician, astrologer, physician and, some say, thief. One of the things he stole was a partial solution to the general cubic equation, which, to give him his due, he improved. His treatment of cubics and quartics, as well as his generous attitude toward imaginary and negative numbers, appear in his *Artis magnae sive de regulis algebraicis liber unus* of 1545. Twenty-five years later he spent a few months in the prisons of the Inquisition, not for his attitude toward negative numbers but for casting a horoscope of Christ.

Ambroise Paré was an army surgeon, hence ignorant of the Latin, and also of much of the nonsense that made up the medical education of the sixteenth cen-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Shock of the new — the Leyden jar, an early type of capacitor invented in 1745. It was used, among other things, to cure paralysis, flatulence and arrogance.

tury. Once while patching up the wounded he ran out of the boiling oil that, in the standard practice of the day, was poured on gun-shot punctures to destroy their "poison". As a makeshift, Paré used a salve of egg yolks and turpentine. Those lucky enough to receive the salve made a faster recovery than their fellows whose injuries had been boiled. Paré experimented with several sorts of unguents and reported his results in his *Méthode de traiter les playes faites par hacquebutes* in 1545. His procedure was later bettered by Sir Kenelm Digby, who removed the salve altogether, or, rather, put it where it would do no harm, on the weapon rather than on the wound.

1495 (5.0 centenary)

We continue in the medical mode. During the siege of Naples by a French army from late 1493, a new form of disease broke out, which by 1495 could no longer be neglected. The French called it the Italian disease and the Italians the French disease. As is often the case in science, they may both have been wrong. Many used to think that it was a pox from the New World, brought to Naples by a friend of a sailor on Columbus's first voyages. In 1530 an Italian astronomer and physician, Girolamo Fracastoro, wrote a poem about a shepherd who suffered from the disease. Fracastoro called the shepherd and his complaint Syphilis, and, of course, attributed it to the French, the full title of the poem being *Syphilis, sive morbus gallicus*. We do not know whether it was new or old, or French or Italian. It was, and remains, *le mal de celui qui l'a*.

1945 (0.5 centenary)

Centennial observances are usually celebratory. The fiftieth anniversary of the bombing of Hiroshima and Nagasaki will be cause for reflection on how humans come to invest their brains and treasure in destroying one another, not for congratulations about the role of science and technology in ending the Second World War. Those who want to remember the achievements of military technology might direct their celebratory impulse to the civilian availability of penicillin as a consequence of its wide deployment during the war, for which, in 1945, Sir Alexander Fleming, Sir Howard Florey and Ernest Chain won the Nobel prize in medicine.

The incoherence of modern life has prevented us from finding the common denominator among the introduction of fluoridation of the water supply, the reporting of the haematopoietic properties of folic acid, the award of the Nobel prize in physics to Wolfgang Pauli for his exclusion principle and the discovery, by Salvador Luria and Alfred Day Hershey, of the mutability of bacteriophages — except that all occurred in 1945. So did the deaths of Hans Wilhelm Geiger, of the counter;

Sir Ambrose Fleming, of the valve; Thomas Hunt Morgan, of the fly; W. B. Cannon, of homeostasis; Robert Hutchings Goddard, American rocket pioneer (compare 1895); and the Russian geochemist Vladimir Ivanovich Vernadsky.

1920 (0.75 centenary)

The matter referred to under 1895, in connection with Grace Young's PhD, continued its slow progress with the very belated admission of women to degrees at Oxford. Knowledge of asteroids (see 1845) had progressed faster, arriving, in



A pair of green woodpeckers and a great spotted woodpecker battling over a hole in an oak tree. From H. Eliot Howard's *Territory in Bird Life* (1920).

1920, at Walter Baade's detection of the 944th, Hidalgo, whose aphelion lies as far out as Jupiter. Then there was more news of, and noise about, nebulae (see 1845 again), when Heber Doust Curtis, pointing to his data about the frequency of novae in Andromeda, argued that it lies outside our Galaxy, and so faced down Harlow Shapley, who thought it was closer, at a famous shoot-out at the National Academy of Sciences.

Many scientific workers survived the war only to succumb in 1920. Among them were Sir Norman Lockyer, astronomer, publicist and founding editor of *Nature*; Robert Edwin Peary, American polar explorer; Srinivasa Ramanujan, self-taught Indian mathematician; Augusto Righi, Italian physicist, one-time teacher of Marconi (1895); William Gorgas, mastermind of the Panama Canal; and Wilhelm Max Wundt, the effective founder of experimental psychology. As a sober reminder of our global village, we recall that, in 1920, the world was still in the grip of the most devastating epidemic in history, the influenza pandemic.

Birds may be no nicer than humans. That follows from the revelation of

H. Eliot Howard's *Territory in Bird Life*, that birds quarrel much more often over real estate than over females. Which reminds us that, also in 1920, John T. Thompson, retired from the US Army, patented what the world did not need, a submachine gun, which its users affectionately called the Tommy gun.

1970 (0.25 centenary)

A quarter of a century ago, the long-time director of the natural sciences division of the Rockefeller Foundation, Warren Weaver (b.1894, 1¢ to within 1 per cent), was in a reflective mood. He published his autobiography as well as a note on the early history of his favourite term, molecular biology. He had reason to feel smug, for the cutting edge of the science was getting ever sharper. Hamilton Smith and Daniel Nathans provided the enzymatic tools of recombinant DNA and genetic engineering; Howard Temin and David Baltimore were busy laying the final demonstration of reverse transcriptase, thus modifying the science's central dogma; Har Khorana, who had already won a Nobel prize, startled the community with his announcement that he had synthesized an artificial gene; and Choh Haohi synthesized growth hormone, 25 years after he had isolated it. Linus Pauling weighed in with his bestselling account of how massive doses of vitamin C could cure the common cold and much else as well.

In 1945 (0.5¢) humans for the first time retrieved something from the Moon. It was a radar signal. Twenty-five years later, Lunik 17 also went to the Moon and back, landing and returning safely, without a crew. In this the Soviet robots did better than the contemporary abortive third manned lunar mission launched by the United States. But the humans outclassed the robots in thinking for themselves. When the oxygen failed in their command capsule, the would-be Moon men crowded into their lunar lander where they could breathe until the airless capsule carrying it returned safely to Earth.

Closer to the ground, the first Boeing 747s went into transatlantic service while the British-French Concorde and the Soviet Tu-144 first flew at supersonic speeds.

To end on an upbeat note, we record that 25 years ago Stephen Hawking worked out that black holes have a temperature. They sweat elementary particles. Consequently, the final state of the Universe will not be a dull collection of black holes, but a dynamic company of black holes, leptons and photons. □

J. L. Heilbron is in the Office of History of Science and Technology, University of California, Berkeley, California 94720, USA; W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Uprooting the human family tree

Our place in nature, seemingly partly settled last year in Ethiopia, is likely to be looked at again and again. No happy ending is yet in prospect.

In palaeoanthropology, at least, the biggest headlines of 1994 were reserved for the discovery and naming of *Australopithecus ramidus* which, at 4.4 million years old, took on the mantle of earliest-known hominid (White, T. D. *et al. Nature* **371**, 306–312; 1994; WoldeGabriel, G. *et al. Nature* **371**, 330–333; 1994).

The creature widely named as the Missing Link was seen within the community more as cause for quiet satisfaction than wild consternation. There was, indeed, a morphological gap between apes and known hominids that was waiting to be filled, and the hominid from Aramis in Ethiopia fitted the bill nicely. The easy placement of *A. ramidus* in the human family tree was a sign of maturity in a field usually distinguished by discord (Wood, B. *Nature* **371**, 280–281; 1994).

Famous last words? Already, there are signs of disquiet. The next few years could see violent upsets in our understanding of human origins, and there are few clues about how a new consensus will turn out. Fossilization is a matter of chance — discovery hardly less so, and the cautious will say (with reason) that consensus is a consummation more to be wished than expected. But in the parlour-game spirit of seasonal prophecy, one can be forgiven for the casting of a few runes.

Those responsible for *A. ramidus* are at present out in the field, and only the most pessimistic will suppose that they will return empty-handed. They will be on the look-out for pelves and leg bones, as these are absent from the published collection, and would confirm or refute the supposition (from the shape of the basicranium) that *A. ramidus* walked erect. This is especially important, because hominids are erect by definition — and erect gait is probably the only feature that *A. ramidus* shares with other hominids. The discovery in Kenya by Maeve Leakey and colleagues of material of a similar age to *A. ramidus* — and including a leg bone — has been widely reported.

If such material confirms the supposition of erect posture in *A. ramidus*, then so much the better. But if it turns out that *A. ramidus* is erect, but in other respects closer to the chimpanzee lineage, then the Hominidae will have one of its primary supports kicked out from under it.

Comparisons of the fossils from Ethiopia and Kenya will be necessary to establish whether *A. ramidus* is one species or more, to measure the degree of variation in the sample that can be attributed to sexual di-

morphism, and to assess whether the bones represent a hominid. If more than one species is present, some may be hominids, others not — but with creatures as primitive as *A. ramidus*, it will be almost impossible to tell the difference.

Debate over *Australopithecus afarensis* (the previous holder of the earliest-hominid title) offers a guide to the issues *A. ramidus* workers will face in the years to come.

Some continue to maintain that *A. afarensis* is a single species, if a highly dimorphic and variable one (White, T. D. *et al.* **366**, 261–265; Kimbel, W. *et al.* **368**, 449–451; 1994). Others suggest that it represents two or more, each with its own distinct relationship with later forms (see Aiello, L. C. *Nature* **368**, 399–400; 1994).

It is therefore possible that *A. ramidus* is neither an ancestor of humanity, nor of chimpanzees, but the first known representative of a hitherto undiscovered radiation of pre-hominid creatures (the 'ramidopithecines', for want of a better word) which might include the ancestors of both.

This group could be seen as a grade intermediate between the Miocene apes and australopithecines. The solution adopted by White *et al.* — that *A. ramidus* is a hominid — is the right one given present evidence, but clearly its status is provisional.

The Miocene apes that lived between 20 and 10 million years ago provide another model for how the debate might turn out. Although known to be widespread and diverse (Andrews, P. *Nature* **360**, 641–646; 1992), there is no agreement about which of these primitive creatures (if any) lies closest to the combined hominid-chimp lineage than to (say) that of the orang-utan (see for example Moyà Solà, S. & Köhler, M. *Nature* **365**, 543–545; 1993; Martin, L. & Andrews, P. *Nature* **365**, 494; 1993). And with every new discovery producing a new candidate, the debate is likely to get more complicated, rather than less.

Given this unpromising start, here are a few predictions about how our understanding of human origins will look in the year 2000. It is rash, naturally, and should be viewed wholly as seasonal entertainment.

The identity of the Miocene ape closest to the hominid lineage will still be unresolved, although opinion will be edging towards *Ouranopithecus* (also known as *Graecopithecus*) as a promising candidate.

The inclusion of *A. ramidus* within the genus *Australopithecus* is regarded as moot, not least by White and his colleagues. By 2000, *A. ramidus* will have been removed to

a new genus, and regarded as a member of what we have dubbed the ramidopithecines. This group will also include material of a similar age from Kenya, and doubtless other sites in East Africa. But arguments will fester about the number of ramidopithecine species, their degree of sexual dimorphism, and which fossils should belong to what species.

The ramidopithecines will be seen as a paraphyletic group, defined on the basis of erect posture, but including within it the ancestors of australopithecines, *Homo* and *Pan*, though possibly excluding *Gorilla*. Erect posture will thus be seen as a primitive feature that chimpanzees have lost, rather than an advanced feature that hominids (*sensu* 1995) have acquired.

The fate of the Hominidae as a concept is less clear. Either it will fade for want of recognizable, definable character states, or it will re-emerge as a larger group that includes *Pan* as well as *Homo* as its living representatives.

As for *A. afarensis*, this will have dissolved into two or possibly three different species, each of greater or lesser kinship to creatures such as *Australopithecus africanus* (Dart's 'Taung baby'), the 'robust' australopithecus such as Louis Leakey's 'Zinj', and the lineage of *Homo*.

Agreement about the division of the spoils of *A. afarensis*, if such agreement is possible, will precede consensus on which of these lies closest to the ramidopithecine ancestral root.

At this point, the debate will impinge on a set of arguments running in parallel about the meaning and status of the genus *Homo* itself. The earliest representative, *Homo habilis*, has already proved somewhat fissile, and some fossils of early *Homo* have been removed to a new species, *Homo rudolfensis* (see Wood, B. *Nature* **355**, 783–790; 1992), including arguably the earliest-known representative of the genus (Schrenk, F. *et al.*, *Nature* **365**, 833–836; 1993).

The referral to *H. habilis* of the diminutive hominid OH62 from Olduvai (Johanson D. C. *et al. Nature* **327**, 205–209; 1987) only heightened long-held suspicions that *H. habilis* may be a conflation of several quite different creatures (Wood, B. *Nature* **327**, 187–188; 1987). More recent work suggests that at least some of these had an inner-ear morphology more like that of apes — or even monkeys — than modern humans (Sporer F. *et al. Nature* **369**, 645–648; 1994). *Ecce Homo?* No doubt there will be an answer, sometime.

Henry Gee

The signature of chaos

Richard V. Jensen

WHAT are the signatures of classical chaos in the quantum world? In recent years surprising answers to this question have emerged in diverse areas from atomic physics to solid-state physics¹. The latest report from the exciting new frontier of quantum chaos² reveals one of the clearest signs of the effects of classical chaos in the quantum resistance of small solid-state devices.

Advances in semiconductor fabrication have made it possible to study the electrical properties of two-dimensional, solid-state devices so small ($\sim 1 \mu\text{m}$) and cold ($\sim 0.05 \text{ K}$) that their length scales are much smaller than the mean free paths for impurity scattering of electrons². In this 'ballistic transport regime' the electrical resistance arises from the elastic scattering of the electrons from the sides of the microstructure, so the conductance properties of these small solid-state devices are strongly dependent on their shapes. For some shapes (such as the stadium shown in Fig. 1a) the classical description of the scattering of the electrons through the device is chaotic. For others (such as the circle in Fig. 1b), the classical ballistic motion is regular. So studies of the quantum mechanical resistance of these devices provide an ideal testing ground for contrasting the quantum properties of systems that are classically chaotic or regular.

One important feature of the quantum

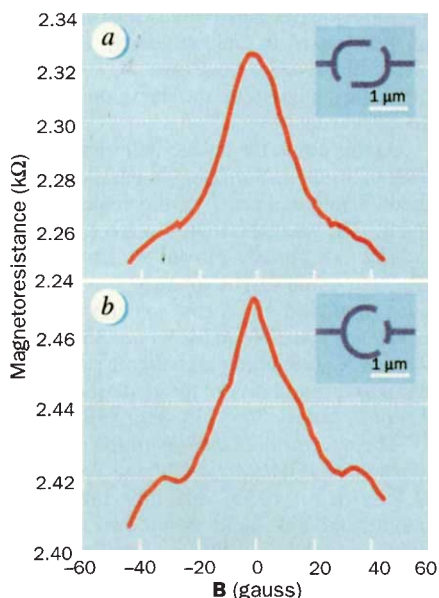


FIG. 1 Magnetoresistance as a function of applied magnetic field, B , for electron current through ultra-cold, two-dimensional semiconductor devices with a boundary shaped like a stadium (a) and a circle (b). In the insets, the thick dark lines represent non-conducting regions. (Adapted from ref. 2.)

conductance is the 'weak localization effect', in which constructive quantum interference raises the quantum resistance above the classical value. The origin of this remarkable effect can be understood using a semiclassical picture of the electron scattering. Here the quantum scattering amplitudes can be expressed in terms of a sum over classical orbits from the entrance to the exit of the device, with each orbit carrying an amplitude and a quantum mechanical phase³. Then time-reversal symmetry alone implies that reflected electron orbits will interfere constructively with their time-reversed pair, enhancing the overall probability of reflection and hence the resistance of the device over their classical values. Magnetic fields introduce an additional quantum mechanical (Aharonov-Bohm) phase to each electron orbit that depends on the 'directed area' enclosed. Consequently, the effect can be destroyed by applying a magnetic field, as the time-reversed pairs of orbits circumscribe areas in opposite directions⁴.

This destruction of the weak localization effect can be observed experimentally by measuring a marked decrease in the electrical resistance of the device as an applied magnetic field is increased. For chaotic scattering domains the 'line shape' of the dependence of the weak localization on applied magnetic field is predicted to be Lorentzian⁴. However, if the classical scattering is regular, general arguments suggest a very different line shape.

The weak localization peaks in the conductance of single ballistic microstructures in the shape of a stadium and a circle have been observed previously⁵⁻⁷. But an innovative experimental approach² was required to observe any differences in the line shapes. The beauty of the latest experiment is that additional quantum mechanical fluctuations in the resistance (the 'universal conductance fluctuations') were cleverly filtered out by simultaneously measuring the conductance of 48 lithographically identical devices fabricated on a single GaAs heterostructure crystal. The results are shown in Fig. 1: as expected, the weak localization peak for the stadium has a Lorentzian line shape with a quadratic maximum at $B = 0$. In contrast, the exceptionally clean data for the circle exhibit a "highly unusual linear line shape"².

This clear difference between the conductance properties of the chaotic and regular scattering domains can be attributed to the striking differences in the area enclosed by typical orbits scattering through the devices, as illustrated in Fig. 2. The regular scattering orbits in the

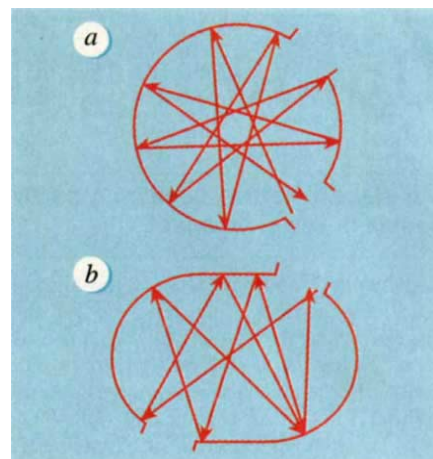


FIG. 2 Typical orbits that bounce 8 times before escaping in a, regular scattering domain; b, chaotic domain. (Figure by W. A. Lin.)

circle are constrained by conservation of angular momentum to circulate in one direction, steadily accumulating enclosed area before escaping, whereas the chaotic orbits in the stadium are constantly winding and unwinding around the enclosed area before escaping. As a result, the typical area and the magnitude of the quantum mechanical phase associated with the magnetic field are expected to be much larger for circles than for stadia of comparable size⁸. This greater sensitivity of the weak localization effect to increasing magnetic field in the circle leads to the sharper decay of the line shape near $B = 0$ shown in Fig. 1b.

Although not yet available in your local video store, this new game of Quantum Pinball⁹ has captured the imaginations of theorists in the field of quantum chaos¹⁰. In addition to the weak localization effect, extensive studies are under way to describe the properties of the large quantum conductance fluctuations in individual devices as a function of applied magnetic field and electron Fermi energy⁵⁻⁷.

These fruitful combinations of innovative experiments and theoretical games do not just promise to improve our understanding of the classical-quantum correspondence for classically chaotic systems. They may also be of practical use in characterizing the unusual quantum mechanical electrical properties of tiny solid-state devices that may be incorporated into the integrated circuits of the future. □

Richard V. Jensen is in the Department of Physics, Wesleyan University, Middletown, Connecticut 06457, USA.

1. Jensen, R. V. *Nature* **355**, 311 (1992).
2. Chang, A. M. et al. *Phys. Rev. Lett.* **73**, 2111-2114 (1994).
3. Jalabert, R. A. et al. *Phys. Rev. Lett.* **65**, 2442 (1990).
4. Baranger, H. U. et al. *Phys. Rev. Lett.* **70**, 3876 (1993).
5. Marcus, C. M. et al. *Phys. Rev. Lett.* **69**, 506 (1992).
6. Berry, M. J. et al. *Phys. Rev.* **B50**, 8857-8860 (1994).
7. Keller, M. W. et al. *Surf. Sci.* **305**, 501-506 (1994).
8. Lin, W. A. et al. *Chaos* **3**, 655-664 (1993).
9. Jensen, R. V. *Chaos* **1**, 101-109 (1991).
10. *Chaos* **3** (spec. issue) 417-706 (1993).

A finger on the culprit

Thomas F. Schulz and Robin A. Weiss

LAST month, to the satisfaction of some and surprise of others, Chang *et al.*¹ announced that they had identified a hitherto unknown human herpes virus in Kaposi's sarcoma. The discovery should provide insights into the aetiology of this unusual neoplasm and more generally into the way in which viruses contribute to the early stages of tumour formation.

First described over 100 years ago², Kaposi's sarcoma (KS) was considered a rare, slowly progressive tumour of elderly men. In the early 1960s it emerged³ that it was endemic in central Africa, where it would sometimes take a more aggressive course. However, KS is now most commonly associated with infection by human immunodeficiency virus, and this aggressive variant⁴, 'epidemic KS', is characterized by rapid progression and a wide distribution which includes the viscera and lymph nodes as well as skin.

Epidemiological information pointed to the involvement of an infectious agent that is preferentially transmitted by sexual contact^{5,6}: KS is common in HIV-positive gay men but virtually absent in haemophiliacs with AIDS; it is also common in HIV-positive female sexual partners of bisexual men but uncommon in intravenous drug users with AIDS. Oral-anal contact might represent a particular risk for transmission of the putative agent⁷, though this effect was not observed in some cohort studies⁸. Kaposi's sarcoma is more common in immunosuppressed transplant recipients than in the population generally, indicating that immunosuppression predisposes to KS and that, if a transmissible agent is involved, it must be present in at least a small population of 'normal' individuals. Prevalence of the agent may also vary geographically, in that, in contrast to Western countries, KS is the commonest post-transplant tumour in Saudi Arabia⁹.

In spite of these strong epidemiological indicators, the search for the agent concerned was long unsuccessful. Herpes-virus-like particles, cytomegalovirus, C-type retroviruses, papillomaviruses and *Mycoplasma penetrans* were considered as possible candidates, each being supported by varying degrees of evidence¹⁰⁻¹².

The herpes virus reported by Chang *et al.*¹ was initially identified using differential representation, a technique based on the polymerase chain reaction that allowed isolation of DNA sequences unique to KS biopsies. The published partial sequence shows some similarity to two known transforming lymphotropic

γ -herpes viruses, Epstein-Barr virus (EBV) and herpes virus saimiri. A region derived from a minor capsid gene can be amplified from virtually all biopsies of AIDS-associated KS and from 15 per cent of non-KS tissue samples from AIDS patients, but not from tissue samples of non-AIDS patients. Our own unpublished data suggest that the virus may also be present in biopsies from HIV-negative post-transplant KS, indicating that, as predicted by epidemiologists, this agent is not confined to HIV-infected individuals.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Kaposi's sarcoma — the ugly reality.

As Chang *et al.* acknowledge, the virus could be a mere 'passenger' in KS tissue; but it is the best bet yet as the long-sought 'KS agent'.

But what exactly is the role of this agent in the development of KS lesions? Why is KS more aggressive in AIDS than in normal individuals or even immunosuppressed transplant recipients? Is KS a fully malignant tumour? Tentative answers to these questions may be emerging from attempts to define early events in the pathogenesis of KS.

Many cell lines established from KS lesions express at least some markers found on endothelial cells and sometimes display fibroblastoid characteristics. They secrete, and depend for their growth, on a series of cytokines¹³⁻¹⁵. For example platelet-derived growth factor and, in particular, basic fibroblast growth factor (bFGF) promote the growth of KS cell lines of an endothelial phenotype in an autocrine manner, and bFGF is thought to be involved in the formation of KS lesions *in vivo*.

The details are controversial, but findings from different laboratories all suggest that several cell lineages (endothelial cells, fibroblasts, epithelial cells) are present in a KS lesion, can adopt a spindle-cell-like morphology, and may depend on each other for certain growth factors acting in a paracrine fashion. Transferred

to nude mice some (but not all) KS cell lines can induce the growth of KS-like lesions of mouse origin¹⁶, implying that factors in these cultures might induce the proliferation of the different cell lineages involved in lesion formation. However, one diploid cell line established from KS in an HIV-negative patient induces KS-like lesions of human origin¹⁷, and at the October Cent Gardes meeting near Paris R. Gallo reported a tetraploid cell line with abnormal karyotype (KS Y-1), established from a patient with AIDS-associated KS, which can cause metastatic lesions in immunodeficient mice. So, although many KS lesions are hyperplastic, rather than neoplastic, neoplastic transformation may occur *in vivo*. Even at this neoplastic stage, however, KS cells may still be subject to hormonal control: at the same meeting Gallo described evidence that β -human chorionic gonadotropin can prevent the induction of KS-lesions by the KS Y-1 cell line.

In a paper that appeared in October, Ensoli and colleagues¹⁸ demonstrated how HIV, which is not an oncogenic virus in the classical sense, may contribute to the initial proliferation of endothelial cells. The Tat protein of HIV-1, when released from cells, can cooperate with bFGF to induce the formation of KS-like lesions in nude mice and enhance proliferation of endothelial cells and KS cell lines. Tat binds to $\alpha_5\beta_1$ and $\alpha_v\beta_3$ integrins by an RGD sequence element in a manner similar to, and replaceable by, their natural ligands fibronectin and vitronectin, and, like the latter, increases attachment of endothelial cells. As bFGF increases the expression of integrins this could explain the cooperative effect of Tat and bFGF, as well as the more aggressive course of KS in HIV-infected individuals.

The newly identified virus may be termed human herpes virus 8 (HHV-8) — or KSV if it does turn out to be causal. At

Sue Ford/SPL

1. Chang, Y. *et al.* *Science* **266**, 1865–1869 (1994).
2. Kaposi, M. *Arch. Dermatol. Syphil.* **4**, 742–749 (1872).
3. Buchbinder, A. & Friedman-Kien, A. E. *Cancer Surveys* **10**, 39–52 (1991).
4. Bayley, A. C. *et al.* *Lancet*, 359–361 (1985).
5. Beral, V., Peterman, T. A., Berkman, R. L. & Jaffe, H. W. *Lancet* **335**, 123–128 (1990).
6. Archibald, C. P. *et al.* *J. acq. Imm. Def. Syndr.* **3**, S18–S23 (1990).
7. Beral, V. *et al.* *Lancet* **339**, 632–635 (1992).
8. Eiford, J., Tindall, B. & Sharkey, T. *Lancet* **339**, 938 (1992).
9. Qunibi, W. *et al.* *Am. J. Med.* **84**, 225–232 (1988).
10. Roth, W. K., Brandstetter, H. & Stürzl, M. *AIDS* **6**, 895–913 (1992).
11. Wang, R. Y. *et al.* *Clin. Infect. Dis.* **17**, 724–729 (1993).
12. Huang, Y. Q. *et al.* *Lancet* **339**, 515–518 (1992).
13. Ensoli, B. *et al.* *Science* **243**, 223–226 (1989).
14. Miles, S. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 4068–4072 (1990).
15. Roth, W. K., Werner, S., Schirren, C. G. & Hofschneider, P. H. *Oncogene* **4**, 483–487 (1989).
16. Salahuddin, S. Z. *et al.* *Science* **242**, 430–433 (1988).
17. Herndier, B. G. *et al.* *AIDS* **8**, 575–581 (1994).
18. Ensoli, B. *et al.* *Nature* **371**, 674–680 (1994).

present, it is unclear whether it infects the proliferating spindle cells which make up the bulk of a KS lesion, or another cell population, thus inducing the proliferation of endothelial cells in an indirect manner, perhaps comparable to that of Tat. The role of the virus in KS could also be similar to that of EBV in some cases of Hodgkin's disease: the persistence of EBV in a latent stage in Reed Sternberg cells may be related to the pathogenesis of this disease, although there remains a possibility that it is a passenger. Finally, in view of the particularly aggressive course of KS in AIDS patients, it may be that HIV and the new virus act additively during the development of KS lesions.

HIV thus appears to make two distinct contributions to the emergence of KS, immune suppression and Tat secretion; yet by itself it is not sufficient to cause KS, as KS seldom occurs in HIV-infected haemophiliacs with AIDS. The new virus may therefore be viewed as a sexually transmitted virus which is relatively common in gay men and is an opportunist in HIV infection. Epidemiological surveys using serology and PCR will determine whether that view is upheld. □

Thomas F. Schulz and Robin A. Weiss are at the Chester Beatty Laboratories, Institute of Cancer Research, Fulham Road, London SW3 6BJ, UK.

CLIMATE CHANGE

Ice sheets and sea level

David Bromwich

THE great ice sheets covering Antarctica and Greenland lock up vast quantities of water extracted from the world's oceans, and so have the potential to affect global sea levels considerably. In the short term, it is the precipitation rate over the ice sheets that determines their varying impact on sea level. On page 52 of this issue, Kapsner *et al.*¹ look at the records of past precipitation and temperature changes from a Greenland ice core, and conclude that temperature changes alone cannot explain the changing rates of snowfall seen in the core. Changes to the dominant storm track, on the other hand, could easily have produced the differences in precipitation.

Quick change

The ice sheets are roughly in balance between ice input from precipitation and ice loss in the form of run-off and iceberg calving from the margins². The exact state cannot yet be determined for either ice sheet. Ice-sheet precipitation removes about 7 mm yr⁻¹ per year from the ocean, and is large in relation to the present 2 mm yr⁻¹ per year of global sea level rise³ (the relative contributions of thermal expansion, melting of valley glaciers, ice sheets, and so on, to sea level rise are uncertain). There is no contradiction here — precipitation rates can change quickly whereas the ice sheets' response to the changed mass input takes much longer.

On timescales of years to decades, the precipitation changes determine the variations of ice-sheet impact, with those over Antarctica apparently dominating⁴. So, when considering the consequences of the steady atmospheric build-up of greenhouse gases such as carbon dioxide, we should also consider what controls precipitation rates over the polar ice sheets.

To examine the controls on precipita-

tion over Greenland for the past 18,000 years as an analogue for the future, Kapsner *et al.*¹ analysed the annual accumulation rate of snow (nearly equal to precipitation) over central Greenland derived from the thickness of annual layers in the GISP2 deep ice core. The time control is particularly good for this interval, so a reasonably accurate annual time series of layer thicknesses can be derived. After correction for the thinning of the layers by the gravity-driven flow of ice, the annual accumulation rate results. The record shows moderate variability within climate epochs such as the Holocene (the past 10,000 years) and the Last Glacial Maximum (more than 15,000 years ago). Rapid transitions of accumulation rate between climate states take place in a few years, particularly from 'cold' to 'warm' states. The transition from warm to cold is not nearly so large or rapid.

Next Kapsner *et al.* considered the relationship between temperature variations and snowfall variations. Temperature is inferred from the stable oxygen isotope composition of the ice ($\delta^{18}\text{O} = (^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}} - 1$), as these are closely correlated for the present climate at high latitudes⁵. Because $\delta^{18}\text{O}$ is actually a consequence of the atmospheric hydrological cycle (evaporation from the ocean, transportation to and recycling along the path to the ice sheet, and the precipitation processes), problems with the correlation appear on close examination⁶. In other words, correlation does not necessarily imply cause and effect. When the climate states change, varying moisture sources become a real and complicating possibility⁷; thus, the oxygen isotope changes are more reflective of source changes than temperature variations. Nevertheless, Kapsner *et al.* convincingly argue that their basic conclu-

sions are not affected by such uncertainties, at least within climate epochs.

According to simple ideas, the limiting factor on snowfall is the amount of water vapour that the air can hold on average. This certainly holds at least in a very broad sense, as today the annual precipitation amounts decrease by a factor of 20 or more from the Equator to the poles in conjunction with the total atmospheric moisture content decreasing by a factor of roughly 15 (ref. 8). But the comparison made by Kapsner *et al.* showed that average temperature variations alone could not explain the accumulation variations seen on a timescale of decades to centuries in the GISP2 core. We have to look elsewhere for the cause of these changes.

Storm tracks

Locally, and for a limited range of precipitation change, the situation is essentially one of moisture turnover rather than the amount of moisture the atmosphere holds on average. After all, precipitation generally takes place only a small fraction of the time and often under atypical conditions of temperature and pressure. Precipitation variations may bear little relationship to the average temperature, as modern observations from many middle-latitude areas readily show⁹. Kapsner *et al.* concluded from their comparison that the explanation for the accumulation changes seen in the GISP2 ice core could lie in changes in the behaviour of cyclones affecting Greenland, manifested by shifts in the dominant storm track that today passes close to the island's southern tip.

This conclusion is consistent with studies of the modern variations of precipitation over Greenland, which, because of limited accumulation data and problems with direct precipitation measurements, must be determined indirectly. For instance, my colleagues and I put daily atmospheric analyses from 1963 to 1988 into a simple model and derived a precipitation distribution that matched the pattern and magnitude of most of the main features in the long-term accumulation record¹⁰. Strikingly, the modelled precipitation showed big year to year increases and decreases that resulted from changes in the position and intensity of the main storm track near Greenland. And, as Kapsner *et al.* found, these changes were not explicable in terms of temperature variations.

Applying these results to the time period studied by Kapsner and colleagues, one finds that it is relatively easy to explain the large accumulation changes between climate states by consistently shifting and changing the intensity of the storm track. Little systematic change within climate epochs would be expected. The between-state changes would probably have been accompanied by dramatic changes in the air temperatures and in the

sea surface temperatures and sea ice conditions in the North Atlantic Ocean.

Hall *et al.*¹¹ last year used a general circulation model to simulate the impact of double the present atmospheric carbon dioxide content on the winter climate of the Northern Hemisphere. In essence, they modelled the global warming problem with an emphasis on the storm track changes. The most marked impact was on the storm track over the North Atlantic Ocean, which was shifted northward and intensified — probably with greater precipitation — in the general area of Greenland, even though only nominal warming was simulated in this region.

The limited accuracy of their model in this area, the concentration on the winter season instead of the summer and autumn when precipitation is at its greatest, and their omission of oceanic changes leave plenty of room for further research and altered conclusions. But this work does provide general support for Kapsner and co-workers' inference that, in a 'greenhouse-warmed world', atmospheric circulation changes may be more important than direct temperature effects in controlling accumulation over Greenland, and thus its impact on global sea level. And, although I have largely been discussing Greenland, storm processes also dominate precipitation formation over Antarctica.

The realization that precipitation variability over the polar ice sheets depends on far more than just temperature changes means that predicting the effects of 'global warming' on sea level is not as straightforward as attempted in the report by the Intergovernmental Panel on Climate Change². As implied by Hall *et al.*¹¹, the precipitation prediction is subtle and involves competing effects. As the spatial variability is high, better predictions will need regional as well as global atmospheric modelling, and many more time series providing representative spatial coverage of Antarctica and Greenland. Glaciological and atmospheric studies need to be integrated, as together they clearly provide far greater insight than either discipline working in isolation. □

David Bromwich is at the Byrd Polar Research Center, The Ohio State University, Columbus, Ohio 43210, USA.

Cold, pain and the brain

Wm D. Willis Jr

THE localization of functions within the brain has been a matter of interest and controversy ever since the debates over phrenology in the early nineteenth century. Modern anatomical, electrophysiological and imaging techniques have allowed the identification of many of the pathways that mediate sensory, motor and cognitive functions. Among the sensory modalities, however, the pathways that mediate pain and thermal sensation remain relatively poorly described. For this reason, the identification of a thalamic nucleus specific for pain and thermal sensation, as described by Craig *et al.* last month¹, represents a notable advance.

Pain and thermal sensations are both conveyed to the brain through the spinothalamic tract, a large bundle of axons running the length of the spinal cord. These axons receive inputs from pain- and temperature-sensitive sensory neurons, and they project to the thalamus, a fore-brain structure involved in relaying sensory information to the cerebral cortex.

The thalamus contains many nuclei specific for various sensory inputs, but the exact pathways for pain and temperature are poorly defined.

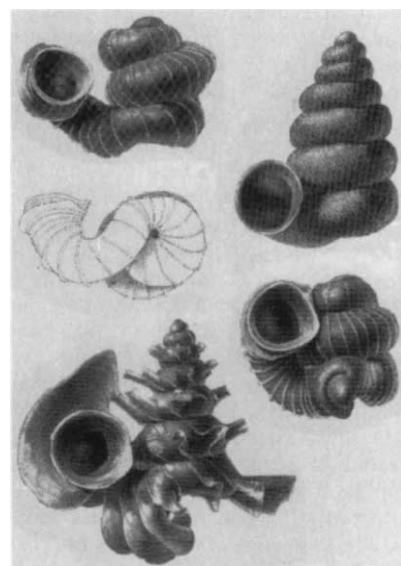
Working with macaque monkeys, Craig *et al.* have identified a nucleus within the thalamus, termed the posterior ventral medial nucleus (VMpo), that seems to be specific for painful and thermal stimuli. They show by anterograde tracing that VMpo is a target for direct axonal projections from lamina I of the dorsal horn, a region of the spinal cord that is responsive to pain and temperature. They also demonstrate that the great majority of neurons within VMpo respond to painful or cold stimuli. The VMpo is known to project to the insular cortex, which in turn has connections with the limbic system, consistent with the idea that this nucleus conveys the affective component of noxious stimuli.

The VMpo nucleus contains a dense plexus of nerve fibres that express the calcium-binding protein calbindin, and

On the one hand ...

THESE snail shells are remarkable snail shells, for they display a hitherto unregarded type of coiling. They are not entirely right-handed (dextral) or left-handed (sinistral), but display elements of both. The specimens are described by J. J. Vermeulen in *Basteria* (58, 73–191; 1994), in a paper dealing with molluscan biodiversity in terrestrial microgastropods of the genus *Opisthostoma* in Borneo.

Thirty-one congeneric species had already been identified, and Vermeulen found a further 36 on the island. In shape and sculpture, their shells are among the most bizarre known. The final whorl is not coiled in line with the preceding whorls but is curved in such a way that the shell has the impression of being sinistral even though the rest of it is dextral. Until now, four kinds of snail-shell coiling have been distinguished, because (although this is rare) the soft body parts can be coiled in the opposite direction to the shell: dextral shells containing sinistral bodies and sinistral shells containing dextral bodies are known. Indeed, things are more complicated still, because the larval shell (which is often retained at the tip of the adult shell) and adult shell can have different hands. R. Robertson has discussed in detail the variations possible in theory, and their occurrence in nature (*Natn. Geogr. Res. Explor.* 9, 120–131; 1993).



J. J. Vermeulen

Vermeulen's specimens provide yet a further twist. One might call their coiling 'sinistroid', to contrast with the term 'pseudosinistral' which is used for a sinistral shell containing a dextral body. But why should this odd form become apparent only now? The reason is that these snails are tiny — the specimens shown here are just 1–3 mm wide.

Edmund Gittenberger

Edmund Gittenberger is at the Institute of Evolutionary and Ecological Sciences, University of Leiden, The Netherlands.

1. Kapsner, W. R., Alley, R. B., Schuman, C. A., Anandakrishnan, S. & Grootes, P. M. *Nature* **373**, 52–54 (1995).
2. Warrick, R. & Oerlemans, J. *Climate Change: The IPCC Scientific Assessment*, 257–281 (Cambridge Univ. Press, 1990).
3. Douglas, B. C. *J. geophys. Res.* **96**, 6981–6992 (1991).
4. Bromwich, D. H. & Robasky, F. M. *Met. atmos. Phys.* **51**, 259–274 (1993).
5. Johnsen, S. J. *et al.* *Tellus B* **41**, 452–468 (1989).
6. Kato, K. *Nature* **272**, 46–48 (1978).
7. Charles, C. D. *et al.* *Science* **263**, 508–511 (1994).
8. Peixoto, J. P. & Oort, A. H. *Physics of Climate* (American Institute of Physics, 1992).
9. Zhao, W. & Khalil, M. A. K. *J. Clim.* **6**, 1232–1236 (1993).
10. Bromwich, D. H. *et al.* *J. Clim.* **6**, 1253–1268 (1993).
11. Hall, N. M. J. *et al.* *J. Q. J. R. met. Soc.* **120**, 1209 (1994).

this distinct staining pattern allowed the authors to identify an equivalent nucleus within the human thalamus. The location of this nucleus is consistent with studies using positron-emission tomography (PET) showing pain responsiveness in the posterior thalamus, and its precise delineation will have considerable implications for the treatment of chronic pain conditions.

The pathway described by Craig *et al.* is clearly not the only one involved in the sensation of pain. Pain has both sensory/discriminatory and affective/motivational components, and these may be mediated by distinct anatomical pathways². In addition to the axons from lamina I, the spinothalamic tract contains many other fibres that originate in other parts of the spinal cord.

For instance, my laboratory has made recordings from thousands of spinothalamic tract cells that project to the main somatosensory nucleus of the thalamus, the ventral posterior lateral nucleus (VPL). Most of these neurons are pain sensitive; unlike the lamina I cells examined by Craig *et al.*, however, none of the cells in our sample responded specifically to thermal stimuli. This suggests that pain and thermal pathways may be distinct, and indeed lesions in the vicinity of the spinothalamic tract can sometimes result in selective loss of either pain or thermal sensations³. On the basis of these observations, Foerster proposed³ almost 70 years ago that the thermal pathway lies dorsal to the pain pathway, and the results of Craig *et al.* support the idea that the thermal pathway may correspond to the lamina I neurons that project to VMpo.

It seems unlikely that VMpo is the only thalamic nucleus involved in pain sensation, for several reasons. Lamina I neurons project to several other thalamic nuclei, including VPL, the ventral posterior inferior nucleus (VPI) and the ventral caudal part of the medial dorsal nucleus (MDvc). Studies with PET have shown that several cortical regions are activated in response to noxious heat stimuli, and some of these regions receive input from VPL (ref. 4). Various groups have demonstrated that neurons in VPL and VPI of monkeys and humans respond to noxious stimuli^{5–7}. Furthermore, microstimulation experiments have shown that these ventral posterior nuclei can mediate pain and temperature sensa-

tions in human subjects⁸. Clearly, these pathways also play an important role in pain sensation.

Because pain has both sensory/discriminatory and affective/motivational components, a reasonable hypothesis is that sensory discrimination is mediated by the pathway from the ventral posterior nuclei to the somatosensory cortex, whereas the affective component that

makes pain unpleasant is conveyed by the medial thalamic nucleus identified by Craig *et al.* and by its connections to the limbic system. □

Wm D. Willis Jr is in the Department of Anatomy and Neurosciences, University of Texas Medical Branch at Galveston, 301 University Boulevard, Galveston, Texas 77555–1069, USA.

TECTONICS

When did India hit Asia?

Rob Butler

MOUNTAIN building resulting from the collision between two continents is among the most dramatic manifestations of plate tectonics. If you want to examine the relationship between large-scale tectonics and the geological consequences of collision, the main playgrounds are the mountain ranges of the Greater Himalaya, Tibet and beyond. In recent years, studies have shown that the mountain ranges have affected not only the immediate environment but also regional climate¹, the chemistry of the world's oceans² and perhaps even the deposition of oil source rocks around the planet³. Mountain building reflects the deformation of the continental crust that followed the collision between India and Asia. Getting the timing of this collision right is

critically important in understanding regional evolution. On page 55 of this issue, Beck and co-workers⁴ report the best estimate yet of the age of collision — and it's up to ten million years earlier than many existing plate reconstructions have adopted⁵.

Up until now, understanding of the timing of collision has come from two distinct sources. One approach uses reconstructions of the past positions of the Indian continent, based on the opening history of the Indian ocean recorded by the magnetic reversal patterns on the sea floor⁵. These records show a wobble in India's motion, followed, 45 million years ago, by a marked deceleration (from about 11 cm yr⁻¹ to 4.5 cm yr⁻¹). Reduced convergence was thought to represent buoyant continental crust blocking the subduction zone, and hence the initial deceleration was used to date collision. The other approach is to look at the geology along the suture between the two continents. Within the Himalayas, vaguely defined Eocene collision ages (about 50 Myr ago) arise from two rather poorly dated rock suites: continental sediments that lie in the suture and the regional change from mantle-sourced, subduction-related granites to those derived from thickened continental crust⁶.

These existing collision estimates are limited by circular arguments. It is a supposition that collision necessarily leads directly to a reduction in the convergence across the plate boundary. There may have been a significant delay as intervening thin continental crust was subducted. Similarly, there may be significant time-lags between



Suture site — high mountains of the Himalayas.

1. Craig, A. D., Bushnell, M. C., Zhang, E.-T. & Blomqvist, A. *Nature* **372**, 770–773 (1994).
2. Gybels, J. M. & Sweet, W. H. *Neurosurgical Treatment of Persistent Pain* (Karger, Basel, 1989).
3. Foerster, O. *Die Leitungsbahnen des Schmerzgefühls und die chirurgische Behandlung der Schmerzzustände* (Urban & Schwarzenberg, Berlin, 1927).
4. Coghill, R. C. *et al.* *J. Neurosci.* **14**, 4095–4108 (1994).
5. Apkarian, A. V. & Shi, T. J. *J. Neurosci.* **14**, 6779 (1994).
6. Casey, K. L. & Morrow, T. J. *Science* **221**, 675–677 (1983).
7. Kenshalo, D. R. *et al.* *J. Neurophysiol.* **43**, 1594–1614 (1980).
8. Lenz, F. A. *et al.* *J. Neurophysiol.* **70**, 200–212 (1993).

collision and the modification of the geochemical reservoirs that are the source of igneous rocks. In Pakistan, for example, mantle-sourced ('I-type') granites continued to be emplaced until about 20 Myr ago⁷, some 20 million years after nearby rocks had not only experienced metamorphism in the Himalayan orogeny but had been exhumed to provide sediment for adjacent basins⁸.

The virtue of studying modern collision belts instead is that Earth scientists can investigate the relationships between various processes and responses within a well-calibrated time frame. It is unwise to use the responses to define tectonic events until processes such as those that control continental deformation or the longevity of geochemical reservoirs are better understood, ideally through the study of modern orogens.

Rather better constraints come from the new work in the western frontier areas of Pakistan, a region more renowned for arms and narcotics smuggling than for its geology. Fieldwork by Beck and his co-workers⁴ integrated with new and published biostratigraphic data is the key. Although the geology is complex, the approach is very simple. They mapped a succession of sediments overlapping the suture between the Indian continent and fragments of old oceanic terranes that constitute the southern flank of the main Asian continent. Unlike the critical sediments in the high Himalayas, these rocks are marine and contain foraminifera from biozone P5 (Upper Palaeocene). This biostratigraphy has been calibrated radiometrically⁹ to give an absolute age of about 57±1 Myr. Collision must pre-date this.

Plate convergence through the Tertiary era was accommodated by subduction of oceanic lithosphere before collision and continental deformation afterwards. Using the supposed 45 Myr age of collision, it has proved possible to match the subsequent plate convergence (about 2,500 km) with continental deformation in Asia⁷. Collision in the Upper Palaeocene increases by 1,000 km the convergence that was accommodated by continental deformation and so reopens the search for suitable structures.

The new age for collision provides a solution to a very worrying part of the deformation story. Metamorphic rocks in the Himalayas of Pakistan have mineral cooling ages of 39±2 Myr (ref. 9). These rocks must have become buried beneath the over-riding Asian continent, heated up and metamorphosed, and then much of the overburden must have eroded away, all in the time between collision and the age of cooling. The problem here lies with the rates of thermal change rather than tectonics. Existing models of heat transfer in rocks¹⁰ make these cooling ages difficult to reconcile with collision later than 50 Myr ago. Putting the collision before 56 Myr ago gives the rocks time to heat up.

But there are broader implications to getting the timing of Himalayan collision right than merely those relating to local geology and tectonics. The collision was a pivotal event in palaeogeography. Before

collision, in Cretaceous times, the proto-Atlantic and Pacific oceans were linked by a great east-west corridor, the Tethys ocean. Reconstructions of Cretaceous palaeoceanography show major west-east circumglobal currents¹¹. It is the India-Asia collision that is most likely to have switched ocean circulation towards the modern arrangement. Getting the timing right will help palaeoceanographers and palaeoclimatologists to make their reconstructions more robust. More basic geological data are needed, probably away from the high Himalayan ranges, to define the collision better and, particularly, the degree of its diachroneity along the length of the old plate boundary. □

Rob Butler is in the Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.

VISUAL PERCEPTION

Seeing backwards in time

Richard L. Gregory

THE phenomenon of backward masking, also called metacontrast, is when a later event influences an earlier event in the visual system. Specifically, a simple stimulus can be erased from perception — that is, it is simply not seen — when followed by a later stimulus presented close to it in space.

On page 66 of this issue¹ Ramachandran and Cobb look into this seeming paradox, and from their results adduce further evidence that top-down processes are able to affect mechanisms of vision that were once thought to be involuntary. The interest here is that perceptual processing is usually thought of as occurring bottom-up from stimuli, running through a cascade of various parallel channels to culminate in behaviour and (mysteriously) awareness. But there is growing evidence that the process is by no means one way, and that top-down mechanisms are involved in interpreting or reading stimulus patterns as more-or-less familiar objects. It has been established anatomically that there are richly diffused descending neural pathways in the brain to support such mechanisms.

Ramachandran and Cobb have carried out experiments to investigate effects on backward masking in subjects who were concentrating on the display pattern (that is, were in a state of visual attention, in trade jargon). The visual display consisted of briefly presented target disks which were immediately followed by flanking squares, shown at different time intervals. The upshot is that by adding more distant visual features to be attended to, Ramachandran and Cobb show that the

occurrence of backward masking can depend on the subject's attentional set, and they suggest that this result demonstrates voluntary top-down modification of the stimulus signals².

These results, and the topic generally, raise several issues. First among them is how a visual process can apparently work backwards in time³. This phenomenon is less of a puzzle when it is realized that, for example, the impression of movement from two actually stationary stimuli stems from processing early in the visual system or even in the retina. This principle is called the phi phenomenon, and in its best known form, a pair of small lights flashed alternately are seen as a single light in motion⁴. What is happening is that the cortex does not receive signals first from one stimulus, then from the other; rather, it receives a single signal from the motion processor.

Backward masking might be closely associated with this early signalling of movement. If it is, as in the phi phenomenon, the first visual stimulus is not being separately signalled from the second. The stimuli would form a single package which has to be unparcelled higher up in the perceptual process — which might interpret the package as arising from two lights (or objects) or from a single, moving light (or single object). Differences in attentional set might affect which interpretation is accepted, much as for other ambiguous visual inputs such as a duck-rabbit or a Necker cube⁵. But this notion of later interpretation of ambiguous initial signals is somewhat different from assuming top-down modification of

1. Prell, W. L. & Kutzbach, J. E. *Nature* **360**, 647–652 (1992).
2. Richter, F. M., Rowley, D. B. & DePaolo, D. J. *Earth planet. Sci. Lett.* **109**, 11–23 (1992).
3. Raymo, M. E. *Palaeoceanography* **9**, 399–404 (1994).
4. Beck, R. A. et al. *Nature* **373**, 55–58 (1995).
5. Dewey, J. F., Cande, S. & Pittman, W. C. III *Ecol. geol. Helv.* **82**, 717–734 (1989).
6. Searle, M. P. et al. *Geol. Soc. Am. Bull.* **98**, 678–701 (1987).
7. George, M. T. et al. *Spec. Publ. Geol. Soc. Lond.* **74**, 173–191 (1993).
8. Treloar, P. J. & Rex, D. C. *Tectonophysics* **180**, 323–349 (1990).
9. Harland, W. B. et al. *A Geologic Time Scale 1989* (Cambridge Univ. Press, 1990).
10. England, P. C. & Thompson, A. B. *Spec. Publ. Geol. Soc. Lond.* **19**, 83–94 (1986).
11. Barron, E. J. & Peterson, W. H. *Science* **244**, 684–686 (1989).

the primary processing or stimuli. Moreover, there may well be more to backward masking than the phi phenomenon. The parallels are not exact in that, for example, phi operates with spatially widely separated stimuli, whereas backward masking only works when successive stimuli almost touch.

It may be helpful to extend the usual terminology of bottom-up signals coming from the senses, and possible top-down effects on them, to include what might be called 'topping-up'; that is, adding knowledge to stimulus signals, which is generally a necessary part of interpreting the ghostly patterns of stimuli as solid objects of the external world. Topping-up with knowledge from prior experience, or with assumptions, may be part of the mechanism for deciding between ambiguous possibilities, while leaving the raw data of the initial neural signals unaffected by what may be false assessments. Thus the perceptual system could continually reassess inputs free from contamination by possibly inappropriate top-down intervention.

Ramachandran and Cobb themselves think of their finding that backward masking is modified by voluntary attention in terms of top-down modification. But the notion of the topping-up of sensory signals by adding knowledge might be worth considering. On this account, attention would select between the possibilities offered by ambiguous signals, rather than modify those signals. Nonetheless, the effects of attention on the visual system are not at all well understood. Ramachandran and Cobb have made a welcome contribution to revealing hidden processes we depend upon, but can also be misled by, for discovering day-to-day reality through our eyes.

An unsolved issue here is the so-called binding problem⁶: how do the various differently processed features of an object cohere, as they are processed in parallel? Should this be thought of as a hardware (squishyware), or as a software processing aspect of brain function? Is 'topping-up' by knowledge of objects, and attention, involved in binding features together to give coherent perception? These murky issues may be illuminated by further investigation of the curious phenomena of perception. □

Richard L. Gregory is in the Department of Psychology, University of Bristol, Woodland Road, Bristol BS8 1TN, UK.

The price of being at the top

R. I. M. Dunbar

THE evolutionary advantages of winning in competitive mating systems have always seemed obvious; and it has seemed equally obvious that the more powerful animals will be in a better position to monopolize whatever resources convey those reproductive advantages. More puzzling has been the question of why such situations do not lead to evolutionary arms races in which individuals become ever more aggressive. The report by Packer and colleagues (page 60 of this issue¹) is of especial interest in this respect because it provides unequivocal evidence for a cost to high rank.

The paper draws on data on 138 female baboons that live alongside Jane Goodall's better-known chimpanzees in Gombe National Park, Tanzania. Packer *et al.* are able to confirm that high-ranking females have shorter inter-birth intervals and higher offspring survival rates than lower-ranking females, as well as offspring that mature earlier (thereby supporting findings from some other long-term studies²). Because fitness is the product of birth rates and the length of the reproductive lifespan, this should imply that females who achieve higher dominance ranks over a lifetime have higher fitness. But analysis of lifetime reproductive success with respect to mean lifetime rank failed to show conventional levels of significance. Packer *et al.* attribute this to a greater incidence of miscarriages among females of higher rank.

Stress-related reproductive failure thus seems to be one counter-selective force preventing female baboons being driven by an arms race into becoming hyper-aggressive (although the need to cooperate in coalitions — of which the social group itself represents but one variety — must be important in social species such as baboons).

The precise mechanism underpinning these high rates of infertility remains unclear, however, and alternative explanations are possible. One is that, although low-ranking females incur significant short-term costs in terms of reduced fertility, they may be able to pursue social strategies that allow them to reduce the impact of these effects over the longer term³. Such strategies might include forming coalitions with adult males who act to buffer females against the stresses they incur from conflicts with other females: the presence of extra males is known to have a beneficial effect on the fertility rates of baboons^{3,4} as well as other species⁵.

A second possible explanation stems from Packer and colleagues' finding that a number of high-ranking females were

chronically infertile. When only females that produced at least one offspring were considered, Packer *et al.* found the expected significant relationship between mean rank and lifetime reproductive output. This prompts the question as to whether the high incidence of sterility in these baboons is related to the epidemic of yaws (a known cause of sterility in humans) that affected the study population some years ago.

A third — and perhaps more interesting — suggestion is that there are two kinds of females who strive to achieve high rank: those whose endocrinology allows them to cope with the stresses they inevitably incur (and who then gain the conventional reproductive benefits of high rank) and those who are less able to withstand the side-effects of intense social stress (and who suffer high costs in terms of infertility as a result). There is evidence that some male baboons exhibit much higher blood concentrations of stress-related hormones than others when competing for high rank⁶; moreover, there may be consistent differences in personality among primates⁷ that could be related to this.

Given that rank does correlate with indices of fitness, a second issue raised by the new paper concerns the mechanism whereby higher ranking females incur their reproductive advantage. Primatologists disagree as to whether it involves stress-related reproductive failure in low-ranking females or such females' lower competitive ability in gaining access to resources. Laboratory studies have adduced considerable experimental evidence in favour of the first claim⁸. With few exceptions^{9,10}, field workers have tended to favour the ecologist's more conventional assumption that food determines everything.

Packer *et al.* show that low-ranking females do not undergo more menstrual cycles before conceiving than higher ranking females, and from this argue that stress cannot be a significant factor in their lower fecundity. By implication, access to high-quality food resources must be the causal mechanism. However, lower ranking females do take longer to resume menstrual cycling following lactational amenorrhoea, which is consistent with experimental evidence that social stress can delay the onset of both puberty and postnatal resumption of menstrual cycling⁸. Moreover, subordinate females, who are wary of the attentions shown to their infants by other group members, may inadvertently delay resumption of menstruation by allowing their infants to suckle more often than less nervous mothers do¹¹ (high rates of nipple contact suppress

1. Ramachandran, V. S. & Cobb, S. *Nature* **373**, 66–68 (1995).

2. Julesz, B. in *Oxford Companion to the Mind* (ed. Gregory, R. L.) 787–793 (Oxford Univ. Press, 1987).

3. Dennett, D. C. *Consciousness Explained* (Little Brown, New York/Allen Lane, London, 1993).

4. Gregory, R. L. *Eye and Brain* 4th edn, 93–117 (Oxford Univ. Press, 1990).

5. Gregory, R. L. *Mind in Science*, 402–415 (Weidenfeld & Nicolson, London, 1981).

6. Crick, F. C. *The Astonishing Hypothesis* (Scribner's, New York, 1994).

the return to menstrual cycling in many mammals, including primates).

Be this as it may, Packer and colleagues' results appear to contradict findings from other baboon populations that, after controlling for climatic effects acting on vegetation productivity, stresses resulting from female-female competition are a key factor accounting for year-to-year variance in birth rates⁹. This raises the possibility that the relative intrusiveness of the two factors varies according to environmental conditions: access to resources will be less critical in richer habitats¹², whereas female-female competition may be less stressful when group sizes are small. It is notable that group sizes were much larger in Smuts and Nicholson's study⁹ (where stress seemed to be the more important factor) than in Packer and co-workers' study (where access to food seems to be the more important).

Considerable advances in understanding primate biology should now begin to

emerge from detailed comparisons between the various long-term field studies, several of which (like that at Gombe) are in their third decade. The financial cost of these studies has been relatively trivial; the benefits will be enormous. □

R. I. M. Dunbar is in the Department of Psychology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

1. Packer, C., Collins, D. A., Sindimwo, A. & Goodall, J. *Nature* **373**, 60–63 (1995).
2. Altmann, J., Hausfater, G. & Altmann, S. A. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 403–418 (Chicago Univ. Press, 1988).
3. Dunbar, R. *Reproductive Decisions* (Princeton Univ. Press, 1984).
4. Dunbar, R. & Sharman, M. *Behav. Ecol. Sociobiol.* **13**, 157–159 (1983).
5. Silk, J. *Am. J. Primatol.* **14**, 111–124 (1988).
6. Salpolsky, R. *Hormones Behav.* **16**, 279–292 (1982).
7. Stevenson-Hinde, J. in *Primate Social Relationships* (ed. Hinde, R. A.) 28–35 (Blackwell Scientific, Oxford, 1983).
8. Abbott, D. H. *J. Zool., Lond.* **213**, 455–470 (1987).
9. Smuts, B. & Nicholson, N. *Am. J. Primatol.* **19**, 229–246 (1989).
10. Dunbar, R. *Behav. Ecol. Sociobiol.* **7**, 253–265 (1980).
11. Gomendio, M. thesis (Univ. Cambridge, 1988).
12. Harcourt, A. H. *J. Zool., Lond.* **213**, 471–487 (1987).

SOLAR SYSTEM

Chiron illuminated

S. Alan Stern

DEEP in the outer Solar System lies 2060 Chiron, a hybrid object which is neither a planet nor a comet but something in between. Now, as reported on page 46 of this issue, Elliot *et al.*¹ have succeeded in directly probing Chiron's atmosphere by observing it occulting a bright star. In doing so, Elliot and co-workers have revealed opacity structures in the coma which tell us something new about Chiron's unique, quasi-bound atmosphere^{2,3} and the nature of its surface activity.

One of Chiron's most compelling attributes is that it appears to have come from the Sun's Kuiper belt⁴ (or, more accurately, Kuiper disk). This disk-like reservoir of comets gained prominence in the late 1980s when orbital evolution simulations first began to convince researchers that most of the short-period comets we observe could not have come from a distant, spherical Oort cloud to their present orbits (the original hypothesis). Instead, it seems, they must come from a more flattened and much closer reservoir.

Although the predicted population of about 10^9 comets in the Kuiper disk was beyond the detection capabilities of virtually all ground-based facilities until recently, it was hypothesized that larger, and therefore more easily detectable, bodies might also be embedded in the disk. After several years of ever deeper searches, planetary astronomers were rewarded in 1992 when David Jewitt and Jane Luu discovered a faint red body of 22nd magnitude orbiting in the frontier

beyond Neptune⁵. Based on its distance and magnitude, this object, provisionally designated 1992QB₁, is estimated to have a diameter near 200 km. This is about 40 times the diameter of a typical comet.

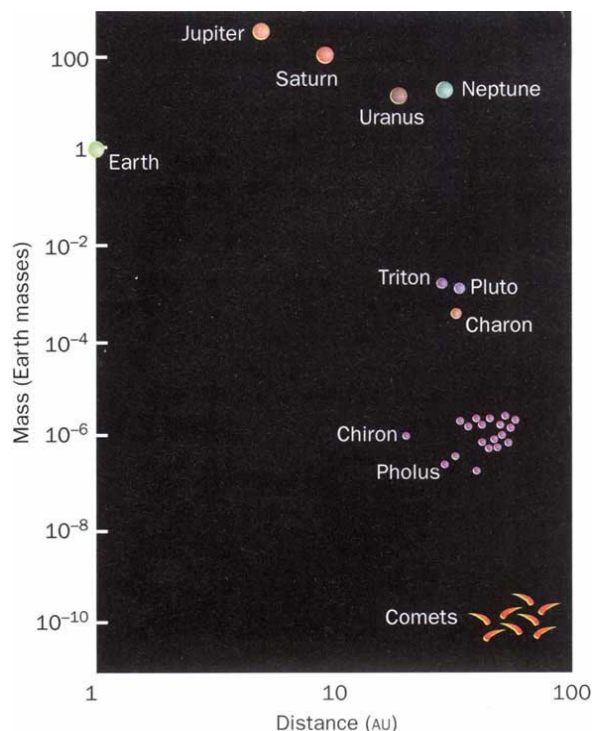
Less than 6 months after 1992QB₁ was found, it was joined by a second object, 1993FW, also discovered by Luu and Jewitt. Although only a tiny fraction of the ecliptic has been surveyed, 17 such objects have now been discovered in the past two years. A simple extrapolation of the existing detection statistics indicates that the Kuiper disk probably contains 20,000–40,000 of these bodies.

Orbits in the Kuiper disk between 35 and 45 astronomical units (1 AU is the distance from Earth to Sun) have been found to exhibit dynamical chaos because they are gravitationally perturbed by the giant planets. When such perturbations occur, the perihelion distance of such objects can rapidly evolve to become Neptune-crossing, and it is then virtually inevitable that they will suffer a close approach to Neptune,

which often results in either a much larger or much smaller orbit. Those objects that are perturbed onto smaller orbits can then be influenced by strong perturbations by the other giant planets. In this way, objects in the Kuiper disk can successively evolve into orbits like those of the short-period comets.

Chiron is among the present population of known outer planet crossers. Although its size has never been determined directly, recent data^{6,7} indicate a diameter between 148 and 208 km. As such, Chiron is 40 times larger and about 6×10^4 times more massive than a typical comet. When it was first discovered⁸ in 1977, Chiron was unique — the Kuiper disk was not then a firm construct and nothing like this 'mini-planet' orbiting between Saturn and Uranus had ever been previously observed.

Chiron's unusual, 51-year orbit ranges in distance from 8.5 to just over 19 AU, with an inclination to the ecliptic plane of 8.5 degrees. Dynamical studies, including the recent work of Dones *et al.*⁹, have shown that this orbit is unstable to perturbations by Jupiter, Saturn and Uranus on timescales of less than 10^6 years. This provides strong evidence that Chiron is a recent addition to the planetary region. Taking this fact together with Chiron's size similarity to 1992QB₁ and its cohorts, as well as evidence for volatiles on its surface (see below), there is a strong circumstantial case that Chiron is an escaped 'QB₁ object' from the Kuiper disk. ►



The architectural outlines of the outer Solar System have been transformed in the past few years by discoveries of numerous Chiron-like bodies with mass near 10^{-6} times that of the Earth. Estimates of the total population of such objects between 30 and 60 AU now range as high as 40,000.

Chiron's activity was first noted¹⁰ when it doubled in brightness in 1988. At the time, it was about 12 AU from the Sun, and the fact that it is active at distances beyond Saturn indicates¹¹ that its source of activity is not driven by water-ice sublimation, but by supervolatile ices such as CO, CH₄ or N₂. Since 1988, Chiron has displayed almost continuous activity in various outbursts lasting weeks to months with amplitudes ranging from several millimagnitudes to over a full magnitude.

Interestingly, images of Chiron have been found on old astronomical plates, taken before its discovery. Analysis of these images¹² has shown that Chiron was intrinsically brighter at its last aphelion in 1970, at 19 AU from the Sun, than in recent years. This distant activity strengthens the case for supervolatiles, but also suggests a complicated variability in the surface activity that may indicate that Chiron's surface is partially or predominantly mantled with inert materials.

The first direct evidence for a coma around Chiron was obtained in charge-coupled-device images² that revealed the presence of small (micrometre-sized) particulates surrounding the object. This particulate coma has since been observed out to distances more than 3×10^5 km from Chiron. Evidence for gas in Chiron's coma was obtained¹³ by spectroscopic observations of the efficiently fluorescing tracer species, CN.

Elliot *et al.* have now extracted fascinating new information about Chiron from photometric studies of the transmission of starlight through Chiron's coma during a rare stellar occultation that occurred in early 1994. The original motivation for observing this occultation was to obtain precise information on Chiron's size and shape, which requires that several stations observe the event. Unfortunately, only two groups, one aboard the NASA Kuiper Airborne Observatory (KAO) and one at the South African Astronomical Observatory (SAAO) obtained data.

With only two lightcurves, a definitive determination of Chiron's precise diameter was not possible. However, something else was found that many will consider much more exciting — the direct detection of opacity structures in the inner few hundred kilometres of Chiron's coma. No such features have ever been reported in any previous stellar occultation by a cometary coma.

Elliot and co-workers' data give clear evidence for no fewer than four discrete 'dust' structures, with optical depths ranging from a few to almost 100 per cent. The number and complexity of the various structures probed in the March 1994 occultation is surprising. And intriguingly, although one of these features ('feature 3') appears to be a broad fan of material, preferentially located in Chiron's orbital plane, features 1, 2 and perhaps 4 are

indicative of particulate arches or jets, presumably emanating from discrete sites on Chiron's nucleus. If this interpretation is correct (additional occultation data are clearly desired), it confirms recent modelling studies¹⁴ that suggest that Chiron's activity, and hence its coma, may be generated by a low-gravity analogue of the geyser-like vents that Voyager 2 observed on Neptune's large satellite Triton.

The Triton analogy, if confirmed, would suggest a possible new connection between Kuiper disk objects and their suspected larger relatives, Triton and Pluto. A particularly important question now on the table is whether Chiron's surface activity and atmosphere is generated by the super-volatile CO, which is common in comets, or by N₂, which is the dominant volatile on Pluto and Triton. If this chemical question can be answered, we will have gained a telling piece of evidence to place the Kuiper disk objects now being discovered in context between the comets and Pluto and Triton.

The observations made by Elliot *et al.* have added impetus to studies of Chiron at a particularly relevant time: Chiron is just now approaching its perihelion, which will be reached in February 1996. Chiron's approach to and passage through perihelion will offer a valuable opportunity to study this fascinating object at four times closer range than its cohorts in the Kuiper disk. The combination of this difference in distance and Chiron's activity makes Chiron appear more than 250 times brighter than QB₁. Perihelion also offers the chance to see this putative Kuiper disk refugee subjected to over an order of magnitude more heating than objects undergo in the Kuiper disk, thereby allowing astronomers to probe its composition, study its surface physics and observe its coma with unprecedented clarity.

Chiron's last perihelion passage occurred in 1945, long before its discovery. The growing body of evidence about this object indicates astronomers will not pass up this opportunity for its successor, in 2047. □

S. Alan Stern is in the Southwest Research Institute, Boulder, Colorado 80302, USA.

1. Elliot, J. L. *et al.* *Nature* **373**, 46–49 (1995).
2. Meech, K. J. & Belton, M. J. S. *Astr. J.* **100**, 1323 (1990).
3. Boice, D. C. *et al.* in *Proc. Distant Comet Activity Workshop* (eds Keller, H. U. & Huebner, W.) 134–139 (Southwest Research Institute, San Antonio, 1993).
4. Kuiper, G. P. in *Astrophysics: A Topical Symposium* (ed. Hynek, J. A.) 357 (McGraw-Hill, New York, 1951).
5. Jewitt, D. & Luu, J. *Nature* **362**, 730 (1993).
6. Sykes, M. V. & Walker, R. G. *Science* **251**, 777 (1991).
7. Campins, H. *et al.* *Astr. J.* (in the press).
8. Kowal, C. T. in *Asteroids* (ed. Gohrels, T.) 436–441 (Univ. Arizona Press, 1979).
9. Dones, L. *et al.* *Bull. Am. astr. Soc.* **26**, 1154 (1994).
10. Tholen, D. J. *et al.* *IAU Circ.* no. 4554 (1988).
11. Stern, S. A. *Publ. astr. Soc. Pacif.* **101**, 126 (1989).
12. Bus, S. J. *Science* **251**, 774 (1991).
13. Bus, S. J. *et al.* *Proc. Distant Comet Activity Workshop*, (eds Keller, H. U. & Huebner, W.) 41–43 (Southwest Research Institute, San Antonio, 1993).
14. Stern, S. A. *et al.* *Astr. J.* **107**, 765 (1994).

DAEDALUS

Sight unseen

*There was a young man who said 'God
Must think it exceedingly odd
If he finds that this tree
Continues to be
When there's no one about in the Quad.'*

Reply:

*Dear Sir, Your astonishment's odd:
I am always about in the Quad.
And therefore the tree
Will continue to be,
Since observed by,*

Yours faithfully, God.

THESE limericks, at least one of which is by Ronald Knox, discuss the old philosophical notion that objects only exist while you are looking at them. Daedalus now has a video system that works this way. It exploits the fact that only a tiny fraction of our field of view (equivalent to a thumbnail at arm's length) can be seen sharply. Our illusion of wide clear sight is produced by scanning this region of high acuity rapidly around the scene. The process can be followed by bouncing an infrared beam off the eye to detect its changing direction of gaze.

So Daedalus's new video system uses an infrared scanner above the screen to detect the viewer and the direction of his gaze. When no one is looking at the screen, it displays only a rough, blurred image, of the sort conveyed by our peripheral vision. But as soon as a viewer glances at any point on the screen, the scanner detects the fact, and the image at that point is drawn in full detail. No matter how fast the viewer looks round the image, the electronics is faster. Wherever he glances, he will find the image clear and highly resolved. He will have the illusion of a complete and perfect picture.

The prime application, of course, is high-definition interactive cable television. Each set will carry an infrared scanner sending the direction of its viewer's glance back to the transmitter. Instead of giving him the whole big picture, it need only send the bit he is currently looking at. A tiny bandwidth could relay the illusion of a stunningly detailed image. Two viewers could even watch different channels on the same screen: each would receive mosaic fragments from his own programme.

Virtual-reality computing will be greatly enhanced as well. Working in real time, even the fastest computer can only create a crude, jerky cartoon world for its user to play in or interact with. But with infrared scanner feedback, it need only define explicitly the small chunk he is currently scrutinizing; it can sketch the rest. The resulting richly delineated fantasy world would be doubly illusory.

David Jones

Earth's peeling veneer of life

SIR — Unprecedented losses of global biodiversity have helped motivate a rapid increase in studies of species extinction. Many ecologists have begun to address the population-dynamic and demographic causes of extinction¹⁻³; others have estimated rates of species loss⁶⁻⁹. The knowledge gained from these initiatives can presumably improve conservation strategies and galvanize public opinion. Yet studies concentrating on the extinction of individual populations and species may overlook broader and more obvious losses of biodiversity.

Many species become extinct only when persecuted by humans, or when their ecosystems, communities and habitats have been dramatically and continuously degraded beyond possible per-

sistence and restoration. The result is an inevitable time-lag between environmental degradation and extinction that highlights the value of more immediate and more easily estimated indicators of human impact.

The table, prepared with the help of S. Jaward, documents alternative metrics quantifying human influences on global biotas. These metrics of human impact, though crude, paint a sombre picture of declining global biodiversity. More than 11% of the world's terrestrial landscape has been converted to cropland. A further 25% is occupied by pastureland. Not surprisingly, the proportion of Earth's surface covered by forest and woodland has dropped by 9% since 1700. Much of what remains is becoming increasingly frag-

mented, and approximately 40% is either plantation or secondary forest¹⁰. Perhaps the most poignant metric is a catastrophic rate of soil degradation, corresponding to 17% of all vegetated land during only 45 years. Biotic functions have been completely destroyed in 9 million hectares and have been "largely destroyed" (ref. 11) in a further 300 million hectares.

Additional metrics confirm dramatic human influence on natural biotas. World exploitation of marine fishes has increased by 35% since 1979. The exploitation of freshwater fisheries has expanded 85% during the same period. Sixty per cent of the world's main fish stocks may be exploited beyond their ecological or economic optima¹². Humans use 8% of the world's available fresh water each year. Tropical deforestation continues at a rate of 0.9% per annum. Human population size is increasing at the rate of 1.7% per year. There has been a parallel increase in the number of domestic animals.

The inescapable conclusion is that much of the planet's surface, and its veneer of life, has already been destroyed. The growth rates of humans, their domesticated mutualists and prey, their co-evolved pests, and their summed demands for resources illustrate that much of remaining biodiversity is in peril. We have embarked on a trajectory of lost biodiversity whose acceleration and direction can, at most, be controlled but not reversed. Painful as it is to reconcile, our efforts toward conserving existing biodiversity need to be balanced by a greater emphasis predicting, and dealing with, the consequences of its dramatic decline.

Douglas W. Morris

*Centre for Northern Studies,
Department of Biology
and Faculty of Forestry,
Lakehead University,
Thunder Bay, Ontario P7B 5E1, Canada*

ALTERNATIVE METRICS HIGHLIGHTING THE LOSS OF GLOBAL BIODIVERSITY					
Extinction	Taxon	Species extinct 1600-1992	%	Species threatened in 1992	%
	Plants	584	0.3	22137	9
	Vertebrates	229	0.5	2212	5
	Other	256	0.02	1353	0.1
Land use	Type	Area ($\times 10^3$ ha)	% Of total	Year of estimate	
	Cropland	1,477,877	11.3	1987-89	
	Pastureland	3,322,943	25.3	1987-89	
	Forest/woodland	4,095,317	31.2	1987-89	
	Other	4,232,737	32.2	1987-89	
Land classification	Class	Area ($\times 10^3$ ha)	% Of total	Year of estimate	
	Protected areas	834,027	6	1990	
	Degraded lands	1,964,400	15	1945-1990	
	Cleared	646,300-653,500	5	1650-1978	
	Forest/woodland	6,215,000	47	1700	
	Forest/woodland	5,053,000	38	1980	
	Tropical forest	1,884,100	14	1980	
	Tropical forest	1,714,800	13	1990	
Fresh water (1987)	Availability and use	Volume ($\times \text{km}^3$)		% Of total	
	Annual renewable water resources	40,673		100	
	Annual withdrawal by humans	3,240		8	
World fisheries	Type	Average annual catch ($\times 10^3$ tonnes) 1987-1989		Increase since 1977-79 (%)	
	Marine	84,220.3		35	
	Freshwater	13,303.2		85	
	Aquaculture	10,575.8			
Population size (1990)	Species	No. ($\times 10^3$)		Annual increase (%)	
	Humans	5,300,000		1.7	
	Cattle	1,271,279		0.5	
	Sheep and goats	1,716,749		1.3	
	Pigs	845,108		1	
	Equines	118,602		0.7	
	Buffaloes and camels	157,967		1.5	
	Chickens	10,399,000		5.3	

Estimates and calculations occasionally vary among references. Sources: extinction, ref. 8, Table 1; land use, proportions of different habitat types, ref. 11, Table 17.1; protected areas, ref. 11, Table 20.1 (does not include marine and coastal areas; the value is an overestimate because some areas are listed under both "national" and "international" protection systems in the reference); degraded lands, ref. 11, Table 8.1; forest clearing, ref. 13, Table 11-1 (low and high estimates); areas of forest and woodlands in 1700 and 1980, ref. 14, Table 10-1 (these estimates cannot be compared directly with those in ref. 11); tropical forest, ref. 11, Table 8.2; freshwater, ref. 11, Table 22.1; world fisheries, ref. 11, Table 23.4 (annual average 1987-89; aquaculture estimate includes fish and shellfish only); population size, humans, ref. 15, p.811; livestock, ref. 11, Table 18.3 (annual average 1988-90).

- Pimm, S. L., Jones, H. L. & Diamond, J. *Am. Nat.* **132**, 757-785 (1988).
- Tracy, C. R. & George, T. L. *Am. Nat.* **139**, 102-122 (1992).
- Schoener, T. W. & Spiller, D. A. *Am. Nat.* **139**, 1176-1207 (1992).
- Diamond, J. M. & Pimm, S. L. *Am. Nat.* **142**, 1030-1035 (1993).
- Caughley, G. J. *Anim. Ecol.* **63**, 215-244 (1994).
- Diamond, J. M. *Phil. Trans. R. Soc.* **B325**, 469-477 (1989).
- Wilson, E. O. *The Diversity of Life* (Belknap, Cambridge, 1992).
- Smith, F. D. M. *et al. Nature* **364**, 494-496 (1993).
- Smith, F. D. M., May, R. M., Pellew, R., Johnson, T. H. & Walter, K. *Trends Ecol. Evol.* **8**, 375-378 (1993).
- Groom, M. J. & Shumaker, N. *Biotic Interactions and Global Change* (eds. Kareiva, P. M., Kingsolver, J. G. & Huey, R. B.) 24-44 (Sinauer, Sunderland, 1993).
- World Resources 1992-93*, compiled by The World Resources Institute (Oxford Univ. Press, 1992).
- Hilborn, R. & Walters, C. *Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty* (Chapman & Hall, New York, 1992).
- Williams, M. *The Earth as Transformed by Human Action* (eds Turner, B. L. *et al.*), 179-202 (Cambridge Univ. Press and Clark Univ. Cambridge, 1990).
- Richards, J. F. *The Earth as Transformed by Human Action* (eds Turner, B. L. *et al.*), 163-178 (Cambridge Univ. Press and Clark Univ. Cambridge, 1990).
- Tolba, M. K. & El-Kholy, O. *The World Environment 1972-1992* (Chapman & Hall, London, 1992).

No electrostatic sense in snakes

SIR — Vonstille and Stille¹ found that a disembodied rattlesnake rattle creates an electrostatic charge when vibrated, as does snake skin rubbed on the substrate, from which they conjectured that rattlesnake tongue-flicking (tongue scanning) is an electrostatic sense used to locate other rattlesnakes and, it is implied, prey as well. Although they stated that "it will be difficult to test the hypothesis", there is a wide literature refuting their conclusion.

Tongue-flicking and the presence of a rattle in snakes are neither logically nor phylogenetically associated. By far the vast majority of snakes lack rattles, but all snakes and lizards (including quadrupedal species that do not drag their skin on the substrate) exhibit tongue-flicking (both air-flicks and substrate-touches), a pleiomorphic feature of squamate reptiles². Therefore, there is neither a phylogenetic nor necessary functional link between the ability to create an electrostatic charge, either by skin dragging or with a rattle, and the behaviour of tongue-flicking. In contrast, the role of tongue-flicking in stimulating the vomeronasal organs (a nasal chemical sense) is unequivocally established³⁻⁵ and is both necessary and sufficient to explain virtually all behaviours associated with tongue-flicking.

The fact that snake skin lacks "electrical discharge points in the form of hair, feathers, bristles, setae or spinules" is more logically related to the phylogenetic constraint of being a squamate reptile, as opposed to a mammal or a bird, than it is to adaptation for generating electrostatic charges. As Vonstille and Stille noted, generation of such an electrostatic charge with the skin is "in common with all other dry-skinned land animals".

Comparative data show convincingly that the rattlesnake rattle is an adaptation for defensive warning⁶, which contradicts Vonstille and Stille's conclusion. Defensive tail vibration in snakes clearly preceded the evolution of the rattle. Because a rattle enhances the performance of an ancestral behaviour, and functions in a defensive/warning capacity at the present time, it is robustly inferred to be an adaptation for this function⁷.

Vonstille and Stille noted the distinction between air-flicks (tongue-scanning) and substrate-touches in snake tongue-flicking behaviour and suggest that air-flicks function for electrostatic sensation. However, despite their assertion that such air-flicks are a "long-standing puzzle", the distinction between air-flicks and substrate-touches has been clearly related to differences in volatility of the chemicals sensed⁸⁻¹², and therefore such behaviour

is once again related to vomeronasal stimulation.

The electrostatic sensory hypothesis¹ requires that the slender prongs of snakes' forked tongues sense slight bending caused by attraction or repulsion of environmental charges. First, a chemosensory hypothesis for the evolution of forked tongues in snakes and other squamates is now overwhelmingly supported. Second, most squamates lack forked tongues but, nonetheless, frequently air-flick. Third, forked tongues have evolved independently up to four times, and the origin of the forked morphology is independent of both limblessness and tail rattles¹³. In other words, ancestral species were unable to generate electrostatic charges, but evolved the lingual behaviour and morphology said to be an adaptation for the perception of such charges. Fourth, histological and ultrastructural studies of snake tongues have failed to find the sensory organs required by the hypothesis for sensing fine displacement^{14,15}.

Vonstille and Stille referred to a behavioural experiment they performed to test their hypothesis. Their brief description strongly suggests that results were random with respect to the snakes' ability to discern electrostatic charges.

We do not dispute the observation that snake skin and rattlesnake rattles produce electrostatic charges, nor that rattlesnakes often flick their forked tongues into the air. However, available data refute the notion of these as adaptations for the production and perception of electrostatic charges. Such electrostatic charges are likely to be epiphenomena arising from structures evolved for other purposes.

Kurt Schwenk

Department of Ecology
and Evolutionary Biology,
University of Connecticut,
Storrs, Connecticut 06269-3043,
USA

Harry W. Greene

Museum of Vertebrate Zoology and
Department of Integrative Biology,
University of California,
Berkeley,
California 94720-3160, USA

VONSTILLE AND STILLE REPLY — We agree with Schwenk and Greene that lizards and snakes use tongues to aid chemosensation — this combination is enhanced by the attraction of particles to the tongue in forms generating electrostatic charges. Snakes differ from nearly all dry skin animals in having no bristles, hair or other static discharge points; natural selection favours static charge production in snakes. A resting rattlesnake, when disturbed, begins flight, circular-in-place movement or rattling before tongue scanning begins — each behaviour produces electrostatic charges. Tail-rattling in ancestral snakes may have had selective advantages in con-

serving energy and distracting attention away from vital body parts while generating charges. Thus ancestral snakes would have increased chemosensory and tongue-flick acuity to identify and differentiate threats from shadows and neutral features to provide a clear functional link between tail rattling, electrostatic charges and snake tongue behaviour.

But in addition to lizard-like tongue-flicks, snakes use tongue scanning. In arboreal and aquatic snakes, tongue scanning has no chemical function in the absence of chemical gradients typical of trails. We believe that snakes use tongue scanning to detect airborne electrons invading plumes of moisture marking cover. Moist air is electrically conductive and a space charge of electrons remains detectable until diluted by turbulence or lost due to other forces.

W. T. Vonstille

W. T. Stille III

Envi-Sci Center,
PO Box 5051,
Winter Park, Florida 32793, USA

1. Vonstille, W. T. & Stille, W. T. III *Nature* **370**, 184–185 (1994).
2. Schwenk, K. *Brain Behav. Evol.* **41**, 124–137 (1993).
3. Burghardt, G. in *Communication by Chemical Signals* (eds Johnston, J. W. et al.) 241–303 (Appleton-Century-Crofts, New York, 1970).
4. Simon, C. in *Lizard Ecology* (eds Huey, R. B. et al.) 191 (Harvard Univ. Press, Cambridge, 1983).
5. Halpern, M. in *Biology of the Reptilia* vol. 18 (eds Gans, C. & Crews, D.) 423–523 (Univ. Chicago Press, 1992).
6. Greene, H. W. in *Biology of the Reptilia* vol. 16 (eds Gans, C. & Huey, R. B.) 1–152 (Liss, New York, 1988).
7. Greene, H. W. in *Biology of the Pitvipers* (eds Campbell, J. A. & Brodie, E. D. Jr) 107–117 (Selva, Tyler, Texas, 1992).
8. Burghardt, G. in *Chemical Signals. Vertebrates and Aquatic Invertebrates* (eds Müller-Schwarze, D. & Silverstein, R. M.) 275–301 (Plenum, New York, 1980).
9. Halpern, M. & Kuble, J. L. *Trends Neurosci.* **7**, 472–477 (1984).
10. Mason, R. T. in *Biology of the Reptilia* vol. 18 (eds Gans, C. & Crews, D.) 114–228 (Univ. Chicago Press, 1992).
11. Alberts, A. C. & Werner, D. I. *Anim. Behav.* **46**, 197–199 (1993).
12. Schwenk, K. *Trends ecol. Evol.* (in the press).
13. Schwenk, K. *Science* **263**, 1573–1577 (1994).
14. Smith, K. K. & MacKay, K. A. *J. Morph.* **205**, 307–324 (1990).
15. Mao, S.-H. et al. *J. Morph.* **208**, 279–292 (1991).

Correction

There was an error in the table of Scientific Correspondence by E. W. Thornton (*Nature* **372**, 327; 1994). The left-hand column should increase in decades reading downwards. The table should read:

HEALTH COSTS OF PRIMARY ENERGY SOURCES	
External health cost adder (pence per kWh)	Source of primary energy
0.01–0.1	Supply of electricity by nuclear, solar, wind, hydro or gas
0.1–1	Supply of electricity by coal or oil
1–10	Energy conservation by draught proofing or double glazing

Holliday junctions cleaved by Rad1?

SIR — Habraken *et al.*¹ recently presented data on the interaction of purified *Saccharomyces cerevisiae* Rad1 protein with small synthetic Holliday junctions. Using a junction that contains a 12-base-pair (bp) region of homology (X12), endonucleolytic nicking was shown to occur throughout the central region of the DNA substrate, leading to its dissociation into products that co-migrated with a duplex marker (*a* in the figure). These results led the authors to conclude that Rad1 protein has an intrinsic Holliday junction resolution activity. However, recent findings by Bardwell *et al.*² showing that Rad1–Rad10 cuts at junctions between duplex and single-stranded DNA (*c* in the figure), together with observations indicating that the junction X12 exhibits a partially single-stranded character, suggest an alternative explanation for the results of Habraken *et al.*

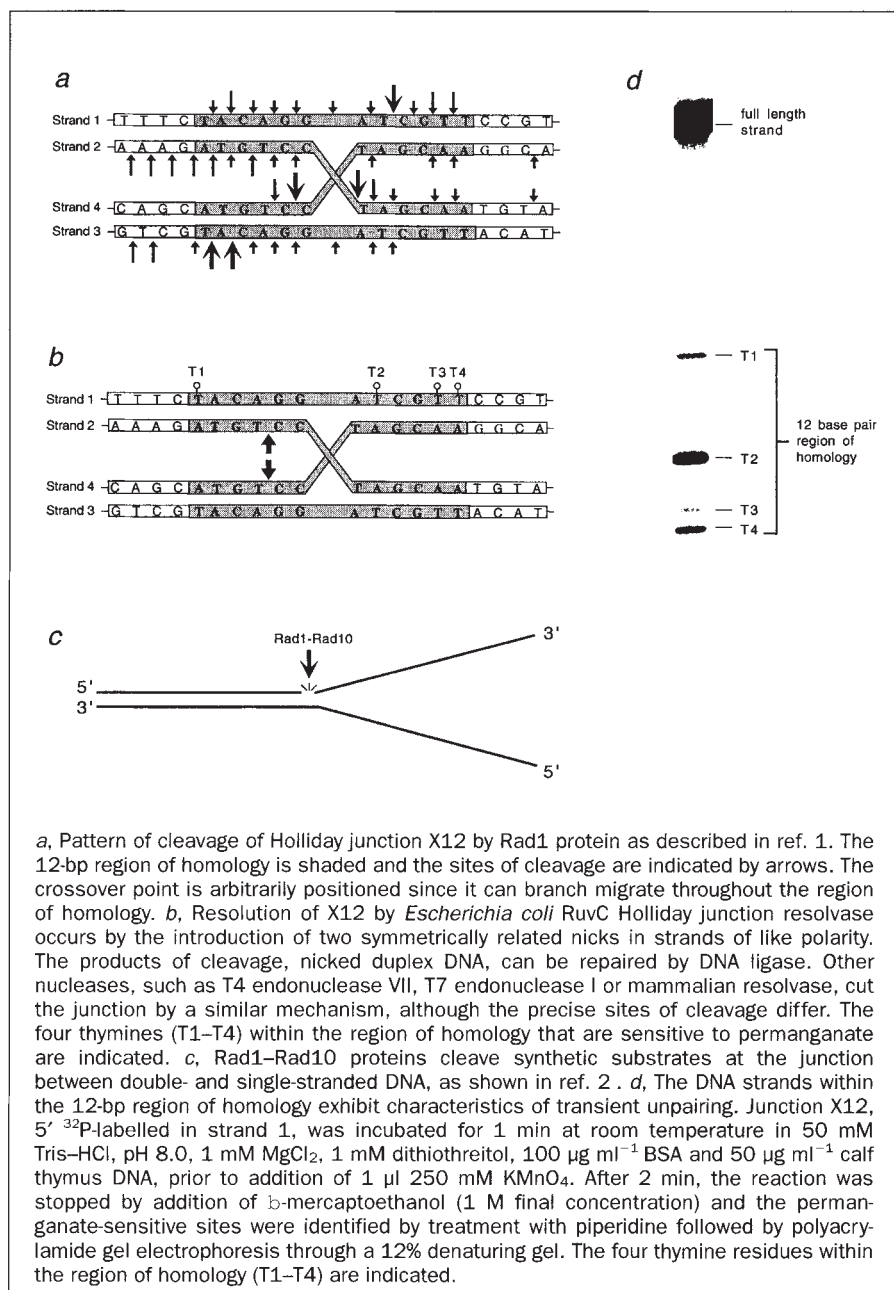
Proteins that resolve Holliday junctions have been isolated from diverse sources including bacteriophage (T4 endonuclease VII, T7 endonuclease I), bacteria (RuvC protein), yeast (CCE1 protein, yeast resolvase) and mammalian cells³. All share the common property of resolving synthetic Holliday junctions (such as X12) by introducing symmetrically related nicks into two strands of like polarity (*b* in the figure). The products of resolution are nicked duplex molecules that can be acted upon by DNA ligase to restore the integrity of the broken polynucleotide chain. T4 endonuclease VII, T7 endonuclease I, RuvC and the mammalian resolvase have also been shown to act upon natural recombination intermediates and promote their resolution into products characteristic of 'patch' and 'splice' genetic recombinants. Properties such as these are required in order to define a protein as a Holliday junction resolvase.

Though Rad1 protein does not cleave the Holliday junction in a manner characteristic of the family of resolvase enzymes, the data of Habraken *et al.*¹ indicate that their Rad1 preparation does indeed catalyse some form of junction-specific nicking. The junction X12, used in the studies of Habraken *et al.*¹, was designed in this laboratory^{4,5}. As a small (~50 bp), easily

prepared DNA substrate that models a Holliday junction, it has served a useful purpose in the study of purified resolvase proteins^{5,6}. However, caution is needed when interpreting results obtained with X12, especially when data are not available with other types of junction substrates (such as natural recombination intermediates). The reason for this is that the duplex DNA within the 12-bp region of homology can exhibit a single-stranded character, indicative of base-pair 'breathing'. When X12 is treated with potassium permanganate, which detects thymine residues in a transiently unpaired state, all four thymines within the homologous region are found to be sensitive to the chemical probe whereas those outside the homologous region are not (*d* in the figure; data provided by R. J. Bennett). Junctions that lack homology, such as X0, which is not

cleaved by Rad1 (ref. 1), do not show extensive sensitivity to permanganate, indicating that the ability to branch migrate tends to destabilize base pairing around the junction point. Thus, the target of Rad1 (ref. 1), rather than the Holliday junction itself (as suggested by Habraken *et al.*), might actually be junctions between duplex and single-stranded DNA, consistent with the results observed by Bardwell *et al.*². Alternatively, the activity reported in ref. 1 may be due to the presence of a contaminating single-strand endonuclease in their Rad1 preparation.

Mutants of *RAD1* and *RAD10* show defects in the excision repair of ultraviolet-damaged DNA. They are also defective in the recombination of repeated sequences^{7–9}, a process that is thought to occur by a mechanism that involves single-strand annealing (SSA)⁹. However,



a, Pattern of cleavage of Holliday junction X12 by Rad1 protein as described in ref. 1. The 12-bp region of homology is shaded and the sites of cleavage are indicated by arrows. The crossover point is arbitrarily positioned since it can branch migrate throughout the region of homology. *b*, Resolution of X12 by *Escherichia coli* RuvC Holliday junction resolvase occurs by the introduction of two symmetrically related nicks in strands of like polarity. The products of cleavage, nicked duplex DNA, can be repaired by DNA ligase. Other nucleases, such as T4 endonuclease VII, T7 endonuclease I or mammalian resolvase, cut the junction by a similar mechanism, although the precise sites of cleavage differ. The four thymines (T1–T4) within the region of homology that are sensitive to permanganate are indicated. *c*, Rad1–Rad10 proteins cleave synthetic substrates at the junction between double- and single-stranded DNA, as shown in ref. 2. *d*, The DNA strands within the 12-bp region of homology exhibit characteristics of transient unpairing. Junction X12, 5' ³²P-labelled in strand 1, was incubated for 1 min at room temperature in 50 mM Tris–HCl, pH 8.0, 1 mM MgCl₂, 1 mM dithiothreitol, 100 μg ml^{−1} BSA and 50 μg ml^{−1} calf thymus DNA, prior to addition of 1 μl 250 mM KMnO₄. After 2 min, the reaction was stopped by addition of β-mercaptoethanol (1 M final concentration) and the permanganate-sensitive sites were identified by treatment with piperidine followed by polyacrylamide gel electrophoresis through a 12% denaturing gel. The four thymine residues within the region of homology (T1–T4) are indicated.

- Habraken, Y. *et al.* *Nature* **371**, 531–534 (1994).
- Bardwell, A.J. *et al.* *Science* **265**, 2082–2085 (1994).
- West, S.C. *A. Rev. Biochem.* **61**, 603–640 (1992).
- Connolly, B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 6063–6067 (1991).
- Bennett, R.J., Dunderdale, H.J. & West, S.C. *Cell* **74**, 1021–1031 (1993).
- Dunderdale, H.J. *et al.* *Nature* **354**, 506–510 (1991).
- Klein, H.L. *Genetics* **120**, 367–377 (1988).
- Schiesti, R.H. & Prakash, S. *Molec. cell. Biol.* **8**, 3619–3626 (1988).
- Fishman-Lobell, J. & Haber, J.E. *Science* **258**, 480–484 (1992).

rad1 and *rad10* mutants do not show defects in other pathways of mitotic recombination, or meiotic recombination, which would be characteristic of a loss of Holliday junction resolvase activity. The role of Rad1 and Rad10 in recombination is more probably related to their function in the processing of double-strand breaks by the removal of unpaired 3' ends⁹.

Stephen C. West

Imperial Cancer Research Fund,
Clare Hall Laboratories,
South Mimms EN6 3LD, UK

Impact modellers not surprised

SIR — Weissman¹ and Chapman² in their News and Views articles made it clear that many astronomers were not expecting to observe effects of the impact of comet Shoemaker–Levy 9 on Jupiter, let alone luminous fireballs and light-scattering plumes rising over the limb. But modellers had predicted that hot debris clouds would rise over Jupiter's limb for sufficiently large impacts^{3–6}.

Chapman² pointed out that there is some variation between the models, and suggested that numerical modellers return to the drawing board because of an apparent lack of evidence from preliminary data from Galileo's photopolarimeter–radiometer for the fireballs that had been predicted. But the half-minute-long light pulses initially interpreted by Chapman as meteor flashes are fully consistent with bolide entry immediately followed by fireball expansion. The channel of hot gas left by the entry radiated the light that Galileo observed, and it is this material in the entry channel that becomes the ballistic fireball. Newer Galileo data and further analysis have now provided strong evidence for this suggestion⁷. We think that Chapman's rejection of the models was premature. Certainly, the next step is to analyse the Shoemaker–Levy 9 impact data and to interpret the results in terms of the computational models. Although the models might need some fine-tuning, they are all in reasonable agreement with many of the observed phenomena.

Mark B. Boslough

David A. Crawford

Experimental Impact Physics Department,
Sandia National Laboratories, Albuquerque,
New Mexico 87185-0821, USA

1. Weissman, P. R. *Nature* **370**, 94–95 (1994).

2. Chapman, C. *Nature* **370**, 245–246 (1994).

3. Zahnle, K. & Mac Low, M.-M. *Icarus* **108**, 1–17 (1994).

4. Stellingwerf, R. F., Hoffman, N. M. & Wingate, C. A. *Bull. Am. astr. Soc.* **26**, 880 (1994).

5. Ahrens T. J., Takata, T., O'Keefe, J. D. & Orton, G. S. *Geophys. Res. Lett.* **21**, 1551–1553 (1994).

6. Boslough, M. B., Crawford, D. A., Robinson, A. C. & Trucano, T. G. *Geophys. Res. Lett.* **21**, 1555–1558 (1994).

7. Weissman, P. R. *Nature* **372**, 404–405 (1994).

Growth limits on phytoplankton

SIR — Morel *et al.*¹ have demonstrated that growth of the planktonic diatom *Thalassiosira weissflogii* decreased in response to zinc limitation. The fact that most of the cellular Zn was found to be associated with carbonic anhydrase, an enzyme central to carbon metabolism, gave rise to the concept of Zn and carbon co-limitation. This, indeed, was supported by the observation that the growth-reducing effect of Zn limitation was partly compensated by increasing CO₂ availability.

Based on these results, Morel *et al.* hypothesized that diminished growth rates under low Zn concentrations reflected a decrease in HCO₃[−] uptake due to reduced carbonic anhydrase activity. To test this hypothesis the authors made use of the equilibrium fractionation of ¹³C between HCO₃[−] and CO₂. Because CO₂ dissolved in water is significantly depleted in ¹³C relative to HCO₃[−] (10.7‰ more negative at T = 10 °C; ref. 2), a decrease in HCO₃[−] uptake relative to CO₂ uptake should result in a more negative organic matter δ¹³C. A lower δ¹³C value in response to Zn limitation was, in fact, observed when Zn-limited cells of *T. weissflogii* were compared with iron-limited cells kept under similarly reduced growth rates. Based on this finding, Morel *et al.* concluded that HCO₃[−] uptake did occur in *T. weissflogii* and was reduced under low Zn concentrations.

Central to the authors' reasoning is the assumption that HCO₃[−] uptake involves carbonic anhydrase (for which, in fact, there is no conclusive evidence). If this assumption were true, however, isotope fractionation by carbonic anhydrase, which for the catalysed conversion of HCO₃[−] to CO₂ is 10.1‰ (ref. 3), would largely cancel the initial difference in δ¹³C of 10.7‰ between HCO₃[−] and CO₂. The δ¹³C signal resulting from carbonic anhydrase-mediated HCO₃[−] uptake and CO₂ uptake would essentially be indistinguishable. Thus, a change in the contribution of carbonic anhydrase-mediated HCO₃[−] uptake in overall carbon acquisition remains undetectable by δ¹³C measurements. (The observed difference in δ¹³C between Zn- and Fe-limited *T. weissflogii* may simply reflect the different metabolic functions of Zn and Fe — Zn in carbon metabolism and Fe in nitrogen metabolism.) In consequence, the results of Morel *et al.* do not permit any conclusion about the source of inorganic carbon acquired by the cell.

Because carbonic anhydrase in *T. weissflogii* is located both extra- and intracellularly, Zn limitation is bound to affect carbonic anhydrase-mediated carbon

uptake as well as assimilation. Variables reflecting the physiological condition of whole cells, such as cell growth, however, do not allow a distinction between intra- and extracellular carbonic anhydrase responses to Zn limitation. The results of Morel *et al.* therefore do not provide conclusive evidence regarding the specific processes influenced by Zn limitation. Although the authors have demonstrated an effect of low Zn availability on carbon metabolism, the origin of this effect remains an open question.

Ulf Riebesell

Dieter Wolf-Gladrow

Alfred Wegener Institute for Polar and
Marine Research,

MOREL AND REINFELDER REPLY — In line with their earlier article⁴ (which showed low growth rates in high pH cultures), Riebesell and Wolf-Gladrow appear to believe that only CO₂ is taken up by marine phytoplankton. As already pointed out by Turpin⁵, however, the uptake of HCO₃[−] by microalgae is not in doubt, and neither is the involvement of carbonic anhydrase in inorganic carbon accumulation (possibly through a simple extracellular catalysis of HCO₃[−]/CO₂ conversion). It would be most surprising if the ability to take up HCO₃[−] were not shared by most oceanic phytoplankton growing at low CO₂.

Nonetheless, Riebesell and Wolf-Gladrow are right to say that high δ¹³C values in phytoplankton organic matter cannot in themselves prove HCO₃[−] uptake, even if many authors have made this inference. Not knowing the exact steps involved in inorganic carbon uptake in marine phytoplankton, we are free to imagine many possible ways other than HCO₃[−] uptake, by which high δ¹³C values can be achieved. Short-term kinetic studies of carbon uptake and characterization and localization of the proteins involved are needed to solve this important problem.

For now, we note that the enhancement of growth rate by Zn in our CO₂-limited cultures implies enhancement of either CO₂ or HCO₃[−] uptake. Based on the literature and our ¹³C and carbonic anhydrase data, we believe the latter to be more likely, but either mechanism would be consistent with our claim that low Zn concentrations may effect CO₂ limitation of phytoplankton in the oceans.

François M. M. Morel

John R. Reinfelder

Department of Geological and
Geophysical Sciences,
Guyot Hall, Princeton University,
Princeton, New Jersey 08544, USA

1. Morel, F. M. M. *et al.* *Nature* **369**, 740–742 (1994).

2. Mook, W. G., Bommerson, J. C. & Staverman, W. H. *Earth planet. Sci. Lett.* **22**, 169–176 (1974).

3. Paneth, P. & O'Leary, M. H. *Biochemistry* **24**, 5143–5147 (1985).

4. Riebesell, U., Wolf-Gladrow, D. A. & Smetacek, V. *Nature* **361**, 249–251 (1993).

5. Turpin, D. H. *Nature* **363**, 678 (1993).

History from a distance

Richard Davenport-Hines

The End of Innocence: Britain in the Time of AIDS. By Simon Garfield. Faber: 1994. Pp. 406. £17.50, \$29.95.

It is too soon to write a definitive history of AIDS, but a historical approach to the epidemic may help in constructing epidemiological models and in studying related policy issues. As exemplified by Mirko Grmek's *History of AIDS* (Princeton University Press, 1990), such an approach can be a valuable contribution to the interdisciplinary study of disease. Grmek's book is purposeful and carefully structured; he is not scared to analyse his material, and he distinguishes between important issues of science and policy and mere ephemera from newspaper cuttings. Unfortunately, Simon Garfield, a British journalist whose previous book was an exposé of the British music industry, lacks Grmek's discrimination and authority.

Garfield opens his history of AIDS in Britain with perhaps its most interesting section: an account of the fatal illness of a young Manchester man, who died in 1959 and whose tissue samples when analysed in 1987 were found to contain the human immunodeficiency virus. Otherwise, his account is familiar: the puzzlement of physicians who first saw AIDS in the early 1980s; the early denial of the existence of a new infectious epidemic; the mobilization of scientific resources to seek treatments and cures; the stigmatization of patients; and the social, economic and political ramifications of the epidemic.

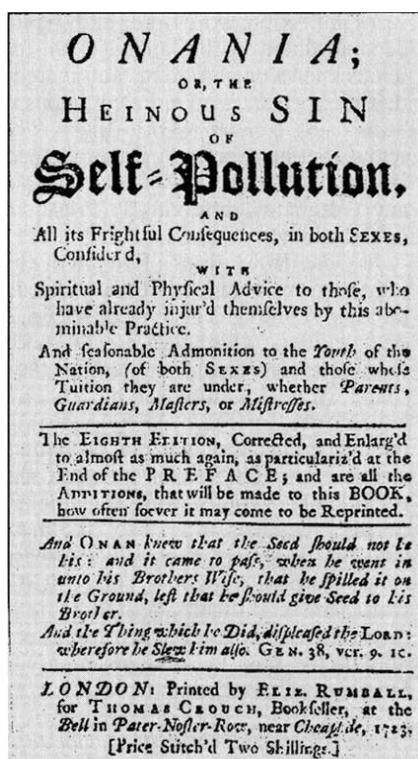
Garfield's publishers stress that he "lives... with his wife and two children", and this commercial decision to distance him from much of the AIDS community in Britain is the source of many weaknesses in the book. He has interviewed politicians, policy-makers, health workers, scientists, patients and carers, and his use of interviews is sometimes powerful, notably in his account of Steve, a young Londoner who became infected with HIV during heterosexual intercourse. There is also a dignified and impressive interview with a young, dying English rock star called Holly Johnson, whose experiences challenge many casual stereotypes of homosexual promiscuity. But, like many journalists, Garfield is easily flattered by the seeming confidences of the great, and too easily swallows their emollient phrases. He is particularly over-receptive to the self-serving account of a former junior government health minister, David Mellor. By contrast, his quotations from his interviews with leading members of the British gay community, such as Simon Watney, seem timid and unimaginative.

The worst omission in Garfield's interviewing is his neglect of the experience of professional nursing of people with AIDS, and his superficial account of the organization of hospice and hospital care. Although he mentions these topics, his quotations from such authorities as Zoe Schramm-Evans or Christopher Spence are so disappointing as to be almost inhumane, and give no sense of their practical experience. Garfield's emphases are on sex, funerals, politics, press stunts, health education programmes, scientific breakthroughs and jealousies; but the most important aspect of life in the AIDS community, caring for sick people, is dwarfed. For this most central aspect of the epidemic, the best source remains Abraham Verghese's magnificent *My Own Country: A Doctor's Story of a Town and its People in the Age of AIDS* (Simon and Schuster, 1994; for a review see *Nature* 370, 606, 1994).

Much of *The End of Innocence* is blandly well-meaning. There is however a serious deterioration in its final chapters, which are an ill-structured morass of journalistic clichés. The penultimate chapter is an account of charity fund-raising that

gives excessive attention to the opinions of television celebrities and other trivialities. Garfield concludes with nearly a hundred pages of extracts from the diary he kept while compiling his book. In a pretentious and meaningless allusion to a historical fiction published by Daniel Defoe in 1722, he entitles this diary "Journal of a Plague Year". Garfield can have learnt little from his investigations if he has not realized how unhelpful to physicians and scientists, how offensive to patients and how inexcusably vulgar is the use of the word 'plague' in the context of AIDS. The diary itself contains some useful data and interviews that would have been better incorporated into the main text of the book, and a good deal of social chit-chat about the Princess of Wales, Hollywood stars and so on. The cursory reference to the attacks on *Nature* for opposing *The Sunday Times's* campaign to prove that AIDS is not a viral disease is so noncommittal that one feels Garfield lacked both the confidence and the time to judge the issues for himself. He does, though, get off the fence for a moment to quote the professor of poetry at Oxford to the effect that although *The Sunday Times* compared its insistence that HIV is not dangerous to its campaign against the thalidomide drug in the 1970s, the truer comparison would be to their purchase in the 1980s of clumsy forgeries masquerading as Adolf Hitler's diaries. □

Richard Davenport-Hines is at 51 Elsham Road, London W14 8HD, UK.



BEING prepared — the anonymous *Onania*, the title page of which is shown here, was the bestselling anti-masturbation tract of the first half of the eighteenth century. Critics argued that such writings would instruct youngsters in a crime they had never even contemplated. Its author's riposte was that it was precisely because self-abuse was such a "heinous sin" that it was crucial that the admonition be broadcast not just to *habitués* but also to adolescents who had "never contracted this guilt". The page is reproduced from the jacket of *The Facts of Life: The Creation of Sexual Knowledge in Britain, 1650–1950* by Roy Porter and Lesley Hall, a detailed and scholarly volume that draws on a wide range of primary sources. To be published on 26 January by Yale University Press at £19.95.

Tongue-tied

Stuart Sutherland

In Other Words: The Science and Psychology of Second Language Acquisition. By Ellen Bialystok and Kenji Hakuta. BasicBooks. Pp. 304. \$27.

RESEARCH on second-language learning has focused on large populations: it has, for example, examined the ease with which Spanish-speaking immigrants to the United States have learned English (or should one say American?). According to Ellen Bialystok and Kenji Hakuta's careful analysis, few firm conclusions have emerged. *In Other Words* does, however, call in question some widely held beliefs. For example, many adherents of Noam Chomsky believe that there is a "critical period" for learning a language during which the innately given linguistic universals, such as the distinction between nouns and verbs, enable the child to make the correct generalizations from the fragments of language heard. Although this may apply to first-language learning, the authors conclude that there is no critical period for learning a second language. Even in studies that show some decline in ease of learning with age, there is no sudden drop at the end of the critical period (which some place at about seven years, others at puberty). Moreover, some well-controlled studies suggest that adults are actually better at learning a second language than are children. Thus do the authors dispel the widely held myth that children have a peculiar talent for acquiring a foreign language.

Scepticism

Bialystok and Hakuta are a highly sceptical pair. They take to pieces the evidence on the personality factors affecting the ease of learning of a foreign language and effectively dispose of them all. It does not matter whether you are an extrovert or an introvert, whether you are an anxious person or whether you have empathy with others. They even, one suspects wrongly, reject intelligence as an important factor by arguing rather lamely that it improves performance on tests of language ability but not on the ability itself.

Their scepticism sometimes produces good knocking copy. They resurrect in a new form the long abandoned Whorf hypothesis that our concepts are to some extent determined by our language: having a term for a concept may focus attention on it. Thai has no terms for 'on' and 'in', but does have terms for 'tight fitting' (a nylon stocking) and 'loose fitting' (my pyjamas). Again they are keen to point out that one of the previously acknowledged language universals (the "complex

noun phrase constraint") does not apply in some languages. It is unlikely that Chomsky will be flying the flag at half mast at the Massachusetts Institute of Technology, for there are many other constraints that are universal. As to the efficacy of the different methods of teaching, none is proven to be better than any of the others. The only exception is that listening to tapes of foreign languages in one's sleep is useless.

One of the few existing findings that the authors accept is that the only aspect of a second language that many never master is pronunciation, although why this should be so remains a mystery. Apparently, however, one of the several recently discovered benefits of alcohol is that in small doses it improves the pronunciation of foreign words.

Omissions

The book has some curious omissions. There is no discussion of people who are exposed from an early age to two languages, for example, children in Wales whose parents converse together in Welsh. On average, such children never become as proficient in either language as do monolinguals in their native language. Nor is there an adequate discussion of the results of learning a language in different ways. Four-year-old children taken to a different country appear to acquire its language in the same way as they learn their native language, in both cases with no conscious knowledge of the syntax that has to be learned. The adult, however, often learns a language laboriously by being systematically taught about nouns, verbs, inflection, word order and so on. To become fluent in the language, this conscious knowledge has to be 'automated' so that the speaker is no more conscious that he is using a noun than a driver is conscious that he is pressing the brake pedal. Do the different ways of learning make any ultimate difference? Presumably nobody knows.

In Other Words does a fine job of clearing the ground. Indeed, it leaves hardly a building standing. This may be useful, but one wonders what edifice will succeed the wholesale demolition. Perhaps the research strategies employed have been the wrong ones. Instead of looking at statistical aggregates, might it not be better to examine the differences between those remarkable people who are fluent in twenty languages and those, like myself, who find it almost impossible to learn a single foreign tongue? This might of course be a difficult method to implement, for the skilled linguist would probably be too busy learning his twenty-first language to have time to be tested. □

Stuart Sutherland is in the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Individual views

John Firor

Global Warming: The Complete Briefing. By John Houghton. Lion Publishing, Peter's Way, Sandy Lane West, Oxford OX4 5HG, UK: 1994. Pp. 192. £16.99 (hbk), £12.99 (pbk).

SIR John Houghton's scientific work and his involvement with the Intergovernmental Panel on Climate Change (IPCC) — a review team instigated by the United Nations — position him well to write a book for nonscientists on the current state of scientific knowledge about a possible human-induced change in climate. His leadership role in the Scientific Assessment Working Group of the IPCC has given him an unusual opportunity to gain an overview of the progress of studies and to sample the range of judgements within the community of climate scientists.

This review of the science of climate change draws heavily on material in the IPCC scientific reports. But the pages of *Global Warming* are also crowded with comments on and discussion of many related topics not found in these reports — some for background information and some to broaden the review. It touches on climates of long ago, the Milankovitch process, chaos in weather forecasting, drought in the African Sahel, global use of fresh water, the Gaia hypothesis, the Rio agreements, engineering advances that could reduce the use of fossil fuels and attempts by economists to estimate the effects on gross national product of climate change.

A guide such as this must inevitably address the alleged virulent controversy within the scientific community that has been highlighted by the popular press. Houghton deals with this challenge by relying heavily on the majority scientific view as represented in the IPCC reports. He gives a fair account of the uncertainties in the science but chooses not to confront critics directly. The most vigorous critic from within the climate science community is smoothly dealt with in a footnote, and Houghton displays, without reference to any controversy, a graph (from the IPCC) effectively countering press reports that a short period of temperature measurements made from a satellite prove that the climate is warming slowly or not at all.

An unusual feature of the book is the author's frank discussion of his Christianity and how it shapes his views about the proper reaction to climate change. Writings on climate change are hardly ever value-free: they are usually in some way influenced by their authors' religious beliefs, economic-system preferences and technological optimism or fatalism. But

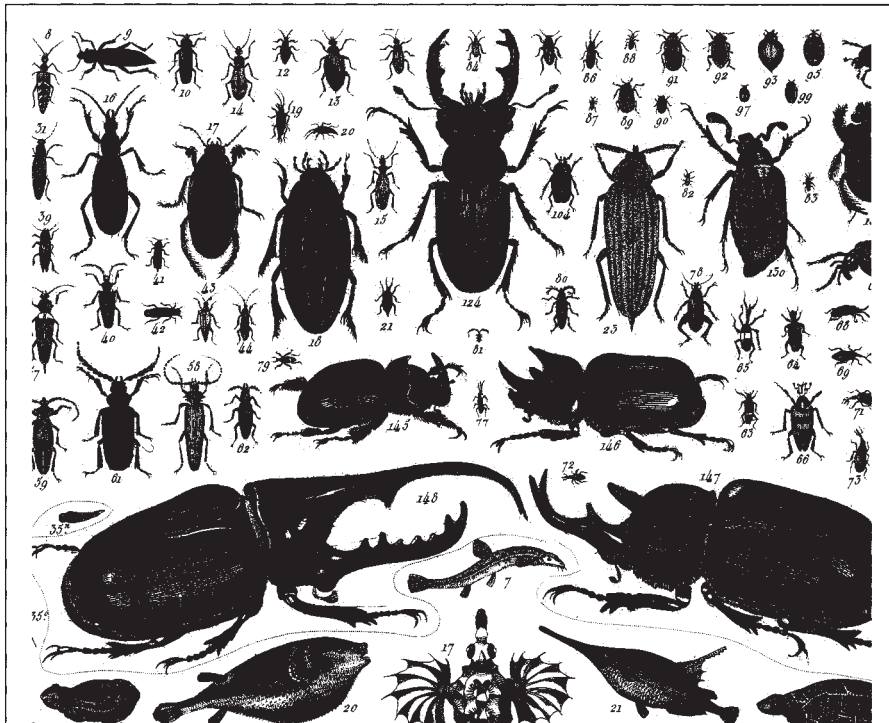
this context is not usually expressly stated, nor is it even clear that the writer is aware that deeply held values can influence scientific opinions. I hope that more authors will adopt Houghton's approach.

The final chapter is also unusual in that the author frames the climate-change problem in the larger array of global environmental issues, most of which are driven by common forces in society. Some of these other global problems have effects that overlap with those of a changing climate: for example, the coastal subsidence we are now causing will have similar effects to a rise in sea level, and a possible mid-continent drying will mimic the overuse of existing water supplies.

The global forces of environmental degradation emphasized by Houghton include poverty, population growth and unsustainable consumption. But, like most reviewers attempting to draw conclusions, Houghton loses heart when he tries to translate these insights into specific actions. Nowhere in his list of "What the individual can do" can be found reactions to the primary forces he has identified — there is no recommendation to limit family size or, for those who live in prosperous countries, to reduce consumption, not only of fossil fuels but of all natural resources. And because depending on the actions of a few knowledgeable volunteers is unlikely to produce appreciable global benefits, he might well have listed the ways in which governments or other organizations could bring about large-scale changes in industrial technology, women's status, economic frameworks and institutional arrangements in order to achieve the needed reduction in resource use and population growth. It should be no harder to recommend the removal of subsidies encouraging the use of fossil fuels or the selling of forests than to recommend that people insulate their homes. Many more people would find their attention engaged and their behaviour changed if such adjustments were achieved. □

John Firor is in the Advanced Study Program, National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307-3000, USA.

■ Three recent challenges to the environmental movement are: **Eco-Sanity: A Common-Sense Guide to Environmentalism** by J. L. Bast, P. J. Hill and R. C. Rue, a concise summary of environmental issues (Madison Books, \$22.95 (hbk), \$12.95 (pbk)); **No Turning Back: Dismantling the Fantasies of Environmental Thinkers** by W. Kaufman, an inside look at how the environmental establishment "ignores basic science, economics and human nature" (BasicBooks, \$25); and **All the Trouble in the World: The Lighter Side of Famine, Pestilence, Destruction and Death** by P. J. O'Rourke, a collection of irreverent and witty essays by an unabashed polemicist (Picador, £14.99).



A SAMPLE of one of 176 beautifully produced sheets in *Heck's Pictorial Archive of Nature and Science*, which presents more than 5,500 copyright-free illustrations of surveying instruments, plant and animal life, minerals, fossils, geological formations, human anatomy and much, much more. The volume is based on the rare original 1851 edition of the monumental *Iconographic Encyclopaedia of Sciences, Literature, and Art* by J. G. Heck, which combined information on a vast range of subjects with some 11,000 steel engravings. Dover, \$14.95 (pbk).

From Czestochowa to Yale

Peter N. Campbell

Eighty Years. By Joseph S. Fruton. *Epikouros*: 1994. Pp.346. Limited edition of 350, available from author.

JOSEPH Fruton and his wife, Sophia Simmonds, known as Topsy, were the authors of the first comprehensive textbook of biochemistry, *General Biochemistry*, published by Wiley in 1953, with a second edition in 1958. At the time this was a very welcome text and did much to establish biochemistry as an academic discipline. In terms of research, with the exception of his war work, nearly all of Fruton's efforts were devoted to problems relating to the biochemistry of peptides, with special reference to the enzymatic hydrolysis and synthesis of peptide bonds.

Fruton was one of the first to direct attention to the mechanism of protein synthesis *in vivo* and I recall that at the Second International Congress of Biochemistry in Paris in 1952 he was a leading light in propounding the role of proteolytic enzymes in this regard. This was, of course, long before the discovery of the genetic code. No matter that his ideas were displaced — he certainly stimulated

many biochemists to come forward with alternative proposals.

Fruton has always been interested in the history of biochemistry and in his later years has written several interesting books on the subject that are characterized by careful and critical attention to detail (for example, *Molecules and Life* and *A Skeptical Biochemist*). Now, at the age of 82, he has published an autobiography. In his preface he modestly excuses himself for this undertaking and mentions his frailties of "arrogance, discourtesy and impatience" that he often exhibited as responses to "like treatment, oversensitivity to personal slight or consequences of resentment, frustration or disappointment".

Fruton was born in Poland, at Czestochowa, in 1912 and was taken by his parents to New York in 1913. In 1917 they left the United States for Minsk, Russia, and as a result of the war the family moved to various countries. His father was employed by Cunard, which sent the family back to New York in 1923.

The description of the struggle of a young Polish Jew to establish himself as a distinguished academic scientist is fascinating. Not only was there the problem of

antisemitism, but also troubles arising from his alleged contacts with communists. For instance, after he had accepted his appointment to the biochemistry study section of the National Institutes of Health he received a letter in July 1952 from the chairman of the board of inquiry on employment loyalty of the Federal Security Agency detailing his contacts. He was asked to complete a notarized "interrogatory" but refused. He was not asked again to serve on any committee responsible to a government agency. For such reasons he does not feel able to describe himself as a patriot of the United States or indeed of any other country.

Fruton describes his time at Columbia and then at the Rockefeller Foundation under the direction of Bergmann. In 1945 he moved to Yale where he spent the rest of his career. He became head of the department of physiological chemistry in 1951, the name of which was soon changed to biochemistry, and retired in 1967, although he stayed on in other capacities. Much of the text is taken up with a blow-by-blow account of the academic politics at Yale. For me this serves to emphasize how much tougher university life is in the United States than in the United Kingdom. Although I, and no doubt others, have had our difficulties as heads of departments, they pale in comparison with those described at Yale. I wonder to what extent the Yale experience is typical — the situation may well have been exacerbated by features of Fruton's character that he describes in the preface and to which I have already alluded.

Other chapters describe the travels of Fruton and his wife in Europe, which often came as a welcome relief from the trials in New Haven. Apart from his descriptions of art treasures to be seen, which demonstrate his catholic taste, and a few intriguing personal comments about his contacts, these accounts are of limited interest. The writing has some amusing touches in that he takes pride in repeatedly stating that "we were taken in a chauffeur-driven car to Heathrow" or some similar phrase. He seems to be unexpectedly interested in zoos and the account of the transfer in 1962 of his Rover 2000 by the SS *Queen Elizabeth* to England and then around Europe is entertaining.

As expected, the text is, so far as I can tell, accurate, which is a tribute considering the age of the author. I noted only two errors: it is Bill, not Bob, Whelan (p.174) and the gas used to whiten flour was agene not agine (p.112); odd that in quite different ways I should have been involved with both of them. □

Peter N. Campbell is in the Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK.

The variety of life

John H. Lawton

Biological Diversity. The Coexistence of Species on Changing Landscapes. By Michael A. Huston. Cambridge University Press: 1994. Pp. 681. £60, \$100 (hbk); £24.95, \$34.95 (pbk).

ALTHOUGH biologists do not know how many kinds of animals and plants there are on Earth, we do know that some places and some habitats hold many more species than others. The difference between the number of tree species in a hectare of tropical forest (often more than 200) compared with a species-rich temperate forest (rarely more than 20) is well known. But many other equally striking patterns exist in the diversity of other taxa along tropical-temperate transects, up mountains, from shallow inshore waters to the ocean depths, along gradients of productivity, and so on. Most of these patterns have been well known for some time, but it is probably true to say that not one of them has a universally agreed explanation.

Huston's book is a bold attempt to find a common link, a grand unifying theory, underpinning many of the major patterns in the way in which organisms assemble themselves into communities on the face of the planet. The core idea is the 'dynamic equilibrium model' of species diversity, which has two elements: the rate at which species are excluded from assemblages by interspecific competition and the frequency and magnitude of mortality-causing disturbances. Local diversity is greatest where neither process dominates. A unique feature of the model is that the same absolute change in the frequency or intensity of disturbance (by fire or hurricanes, for example) can have totally different effects on species diversity depending on the rate of competitive displacement. The model is an amalgam of two previous attempts at a grand synthesis, J. H. Connell's 'intermediate disturbance hypothesis' and J. P. Grime's 'intermediate productivity hypothesis'. To take one of many examples from the book (Fig. 13.1, p.415), Huston suggests that tropical rainforests are so rich in tree species because they have relatively low rates of plant growth compared with temperate hardwood forests (a surprising conclusion but, averaged over the growing season, one that seems to be true) and lower frequencies of disturbance. Boreal forests have fewer species still because they have comparable rates of growth and exclusion to tropical forests but much higher rates of disturbance.

A feature and strength of the text is the huge amount of literature, much of it published decades ago, synthesized and sum-

marized within its 15 chapters. Huston takes the view, with which I agree, that if we are to understand major patterns in the diversity of insects, or birds, or zooplankton, it is sensible to start by trying to understand plant species richness on land, and the diversity of similar structurally important organisms in the sea, such as coral reefs. A key feature of *Biological Diversity* is the distinction it makes between 'structural' taxa that give shape and form to assemblages and 'interstitial' taxa that hitch a ride on, in and under the backs of structurally important trees and so on. The main emphasis of the book is therefore on plant communities on land, although a whole chapter is devoted to marine ecosystems. Huston also distinguishes between the roles of interspecific competition and disturbance that operate primarily within functional groups and the principal consequences for diversity of adding entirely new functional groups, although how one is supposed to identify members of different groups is not made clear.

Does it work? Yes and no. No single, unifying theory is likely to account for all major patterns of species diversity; as Robert MacArthur reminded ecologists more than 20 years ago, we will need contingent theories tailored to taxa and habitats. Huston's is a brave attempt to forge a key element of that emerging contingent theory, and he is undoubtedly correct to emphasize competition and disturbance among structurally dominant taxa. The weakness of the book, which Huston freely admits, is the purely qualitative test of theory; there are no hard numbers on the axes of disturbance and rates of exclusion. A second problem is the weak link between the local processes of competition and disturbance, and regional, landscape levels of diversity; somehow the scales seem ill-matched. A third difficulty I had with the book is Huston's propensity for making grand generalizations, with no supporting evidence, for instance: "maximum diversity of the high trophic levels occurs at levels of productivity higher than those at which lower trophic levels reach their maximum diversity" (p. 538).

It is easy to carp, and to do so is unfair. I prefer to see *Biological Diversity* as full of interesting, untested ideas, underpinned by a unifying theory that now requires rigorous quantitative examination. Therein lies its major challenge to ecologists, faced with understanding the variety and distribution of life on Earth, before human beings obliterate most of it. □

John H. Lawton is in the NERC Centre for Population Biology, Imperial College, Silwood Park, Ascot, SL5 7PY, UK.

■ Just published is *Linking Species and Ecosystems* edited by Clive G. Jones and John H. Lawton. Chapman and Hall, £39.50.

Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS

Kurt Wiesenfeld & Frank Moss

Noise in dynamical systems is usually considered a nuisance. But in certain nonlinear systems, including electronic circuits and biological sensory apparatus, the presence of noise can in fact enhance the detection of weak signals. This phenomenon, called stochastic resonance, may find useful application in physical, technological and biomedical contexts.

SINCE the early days of radio, when the random arrival of electrons at the anode of a vacuum tube translated into an audible and annoying hiss at the loudspeaker, engineers have sought to minimize the effects of noise in electronic circuits and communication systems. But recent research has established that noise can play a constructive role in the detection of weak periodic signals, via a mechanism known as stochastic resonance (SR). In essence, SR is a nonlinear cooperative effect in which a weak periodic stimulus entrains large-scale environmental fluctuations, with the result that the periodic component is greatly enhanced.

Stochastic resonance is now known to occur in a wide range of physical systems; however, it was originally proposed as an explanation of the periodic recurrences of the Earth's ice ages, which exhibit a 100,000-year periodicity. In 1981, a group of European scientists¹⁻³ described a general dynamical mechanism whereby small periodic perturbations could be greatly amplified by large environmental fluctuations. In the case of the ice ages, a weak periodicity in the insolation stemming from variations in the Earth's orbital parameters might cause regular transitions in a bistable energy potential used to model long-term changes in global climate⁴. Although the proposition that the glacial-interglacial cycles are indeed periodic and linked to the Earth's orbital motions has been a matter of recent debate^{5,6}, the possible role of SR in palaeoclimatology is a topic of continuing interest^{7,8}. These discussions notwithstanding, SR as a physical process is now firmly established as a genuine and even common phenomenon.

Experimentally, SR was first demonstrated with a noise-driven electronic circuit known as a Schmitt trigger⁹; this work was also the first to characterize the phenomenon in terms of a signal-to-noise ratio. It took five more years before the interest of physicists ignited, sparked by the demonstration of SR in a bistable ring-laser experiment¹⁰. Stochastic resonance has been reported in a wide variety of physical systems^{11,12}, and the classical theory¹²⁻²⁰ is well in hand. Now SR has crossed disciplinary boundaries: its role in sensory biology is being explored in experiments on single crayfish neurons, and in perceptive brain function by experiments on people's ability to resolve ambiguous figures. These new efforts, together with attempts to exploit SR for technological advantage, are the main trends in current research on this topic. There are even indications that SR may affect research in medical and environmental science.

The mechanism

The basic picture of SR can be illustrated using a mechanical analogy. Imagine a particle subject to friction, moving in a double-well potential. A weak signal serves to periodically modulate the potential by alternately raising and lowering the wells relative to the barrier (Fig. 1a). Here, "weak" means that the modulation is far too small to excite the particle over the barrier. On the other hand, the presence of random noise alone is sufficient to induce (irregular) switching between the wells. In

the high-friction limit, the dynamics can be modelled by the differential equation

$$\frac{dx}{dt} = -\frac{dU}{dx} + F(t) + A \sin(\omega t)$$

where U is the bare potential, $A \sin(\omega t)$ is the signal, and F is the noise. Stochastic resonance is a nonlinear cooperative effect whereby the small signal entrains the noise-induced hopping, so the transitions are surprisingly regular. What is more, the regularity can improve with the addition of more noise. In this way a small regular influence can have a large effect if environmental fluctuations are available to be tapped.

There is another way to view SR which is enlightening when thinking about applications, both in technology and biology. It concerns the problem of detecting weak signals in a noisy environment. Imagine that the system and signal are hidden from view, and that we gain information only by observing the system's output as a time series of switching events. In SR the system becomes a more sensitive detector as more noise is added, at least up to a point: it is optimally sensitive at some non-zero level of input noise.

Today, we know that SR is even more general than the bistable picture implies. Even simpler systems²¹, including those with a single potential well²² and integrate-and-fire²³ dynamics can exhibit SR or SR-like properties. The latter is a common model for neurons wherein a steady input is integrated until the result exceeds a critical threshold whereupon the neuron 'fires', and resets the 'integrator' to zero. In fact, the simplest possible system²⁴ consists only of a threshold, a subthreshold signal and added noise, as depicted in Fig. 1b. Whenever the noise plus the signal crosses the threshold in one direction, it triggers a narrow pulse in the output (Fig. 1c). The nonlinearity in this system is simply the on/off nature of the output. This version of SR has obvious appeal for those working in sensory biology, as neurons are excitable systems with properties similar to those depicted in Fig. 1b, c: when some internal threshold is crossed the neuron 'fires' and then resets itself to await another threshold-crossing event.

Stochastic resonance should not be confused with 'dither', also known as stochastic linearization, a technique wherein periodic or random forcing is intentionally introduced to overcome regions of 'dead' dynamical behaviour in self-regulating systems. A familiar example is the use of dither to cancel the effect of gear backlash which, in servo mechanisms, often leads to undesirable effects such as gear chatter.

Quantifying SR

The most common way to quantify SR is through the signal-to-noise ratio (SNR). This is readily obtained from the output by forming the power spectrum, which measures the frequency content of a time series. A typical result is shown in Fig. 1d, taken from numerical simulations of a threshold-crossing system. The signal shows up as a sharp peak located at the signal frequency riding on a broadband noise background; the SNR is the ratio

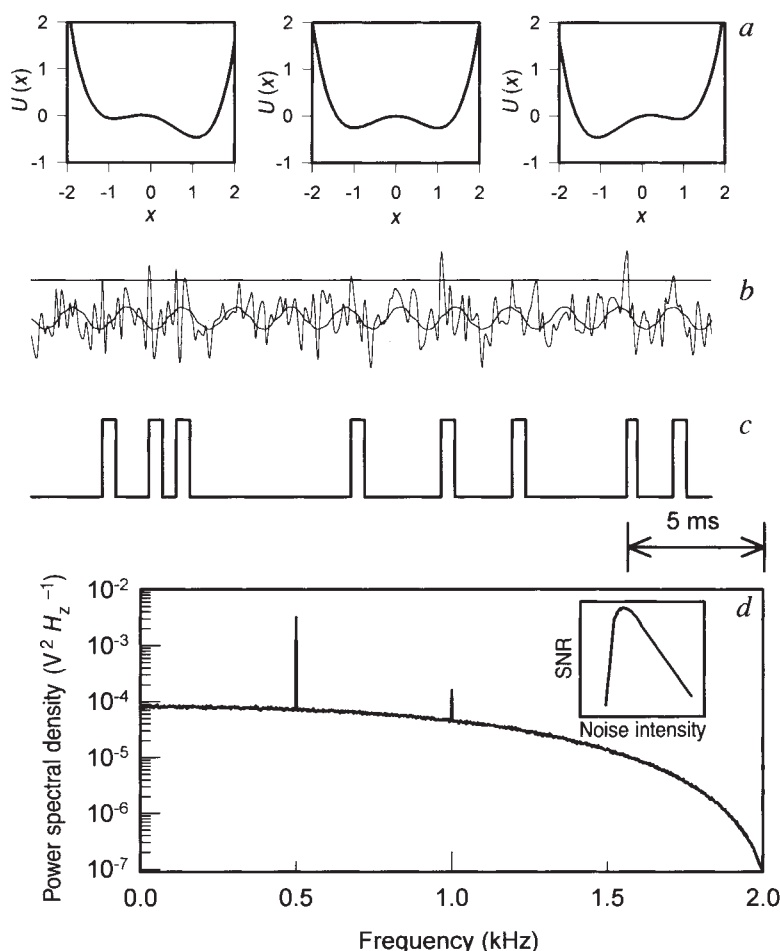


FIG. 1 *a*, A bistable potential, weakly modulated by a periodic signal. *b*, The simplest instance of stochastic resonance (SR) consists of a threshold (shown by the straight line) and a subthreshold sinusoidal signal with added gaussian noise. Each time the signal plus the noise increases across the threshold, a pulse of standard shape is written to the time series shown in *c*. *d*, Power spectrum of a long time series of the pulses shown in *c*. A broadband noise background whose amplitude at zero frequency is related to the mean pulse repetition rate and hence to the noise. The signal feature is the sharp peak located at the signal frequency (0.5 kHz in this case) riding on the noise background. The SNR is defined as $10 \log_{10}(S/N_0)$ in decibels (dB) where S is the area enclosed above the noise background, and N_0 is the amplitude of the noise background at the signal frequency. Inset, the characteristic signature of SR, that is, a maximum in the SNR at an optimal value of input noise intensity.

of the strength of this peak to the background level (see Fig. 1 legend). Remarkably, theories for all three main types of SR—the bistable potential model, the fire-and-reset excitable system model, and the simple threshold model—result in the same general formula (apart from some constant factors of order one in both the prefactor and the exponent) for the SNR:

$$\text{SNR} \propto \left(\frac{\varepsilon \Delta U}{D} \right)^2 e^{-\Delta U/D} \quad (1)$$

where ε is the input signal strength, D is the input noise intensity and ΔU is a constant related to the barrier height or the threshold. The signature of stochastic resonance is that this SNR is zero for zero added noise (that is, $D \rightarrow 0$: no noise implies no switching or threshold crossings, hence no output), rises sharply to a maximum at some optimal noise intensity, and decreases gradually for larger noise intensity as randomization overrides the cooperative effect. The detailed shape of this curve depends on the signal frequency and other system parameters; Figs 3 and 4 show typical examples.

Although the power spectrum is the most widely used coherence measure, it is not the only possibility. An alternative is the residence-time probability distribution²⁵ depicted in Fig. 2*a–c*, more familiarly known to experimental biologists as the interspike interval histogram²⁶. This measure is composed of a set of peaks which are widely spread for very small noise but become more coherent for larger noise intensity. This qualitative structure of the histogram is very common, and is not in itself a signature of SR. Rather, SR is identified with the more particular behaviour shown in Fig. 2*d*, where the amplitudes of the lower-order peaks pass through individual maxima with increasing noise intensity^{25,27}. Similar histograms from both neuron models²⁷ and actual biological preparations²⁸ show SR or at least a strong connection between the ability of a sensory neuron to transmit coherent information and its internal or external noise^{29,30}.

SR in biology

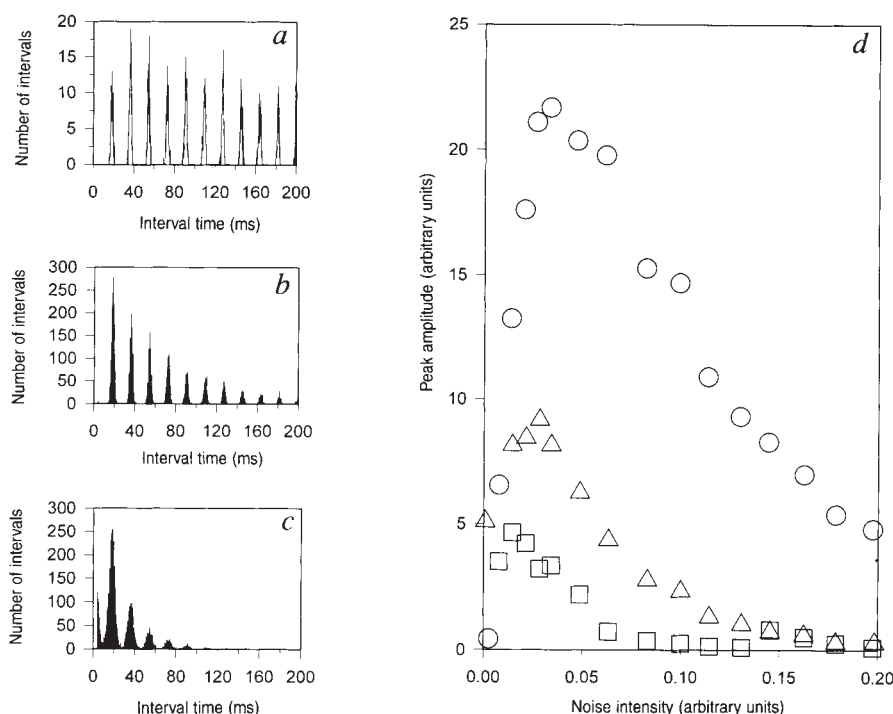
Is it possible that noise-mediated detection or transmission of weak signals plays a significant role in biology? Sensory neurons are notoriously noisy, and in their coarse behaviour operate as threshold systems. Could SR account for the exquisite sensitivity of some animals to weak coherent signals embedded in a noisy environment? This question prompted a series of experiments with a simple, perhaps even primitive, sensory system: the mechanoreceptor hair cells of the crayfish *Procambarus clarkii*²⁸. These cells are specialized to detect weak, coherent water motions produced, for example, by the approach of a predator, perhaps a swimming fish. In these experiments, a piece of the telson together with the nerve cord and the sixth ganglion was excised and mounted in crayfish saline on an electromechanical transducer. Signal and noise generators drove the transducer so that the hair cell's motion through the saline was a combination of coherent and random motions. Electrophysiological recordings were made from a single cell stimulated with a weak (subthreshold) signal. Most cells tested showed clear evidence of SR (Fig. 3).

A familiar and frequently used theoretical model of a neuron, known as the Fitzhugh–Nagumo equations, gave similar results. These were implemented using an analogue electronic circuit operating in the subthreshold regime and stimulated with gaussian noise and a weak sinusoidal signal³¹. Figure 3 shows data from

both the crayfish experiment (squares) and the Fitzhugh–Nagumo simulation (diamonds); both data sets show the characteristic signature of SR. The crayfish data do not decrease rapidly for small noise because of the residual internal noise of the neuron. The solid curve shown in Fig. 3 is a fit to the data using equation (1). Thus theory, simulation and experiment all provide support for the notion of SR as a viable mechanism in sensory biology.

It is possible, although not yet demonstrated, that SR is a common phenomenon in sensory biology. Virtually all sensory systems operate as threshold detectors, and the crayfish mechanoreceptor example demonstrates that the efficiency for detecting weak signals can be enhanced by the addition of external noise. This is obviously helpful in noisy environments, but might even be useful where external noise is not available. Neurons can also have substantial amounts of internally generated noise, and it is natural to ask whether this apparently undesirable noise source can serve a useful function. This remains an important open question: to date, experiments designed to study the role of the internal noise have been inconclusive³⁰.

FIG. 2 a–c, Interspike interval histograms, as measured on an electronic Fitzhugh–Nagumo neuron model³¹ for increasing noise intensity. The area under the first peak represents the total number of action-potential spikes that occur at every cycle of the periodic stimulus. The areas under the higher-order peaks represent events that skipped stimulus cycles. The tendency of the noise to concentrate area under the first peak represents increasing coherence. d, Behaviour of the amplitudes of the higher-order peaks with increasing noise intensity for a bistable system²⁵. The circles, triangles and squares represent data for the second-, third- and fourth-order peaks, respectively.



Technological applications

Researchers are just beginning to pursue possibilities for technological exploitation of SR. With a view to possible applications in magnetic sensing, a group associated with Quantum Magnet-ics, Inc. (San Diego) demonstrated SR in a bistable supercon-ducting quantum interference device (SQUID) loop^{32,33}. Future directions for this work can be threefold. First, the passage of individual fluxons into and out of the loop are macroscopic quantum tunnelling events and thus are candidates to demon-strate quantum SR. Investigation of these events would be of fundamental interest; recent theoretical work³⁴ predicts subtle differences from the classical case with respect to underlying symmetries of the system. Second, SQUIDs fabricated from high-temperature superconducting materials are in principle much cheaper to operate, but are inherently noisy. It may be possible to optimize their performance in some applications by using SR to exploit the internal noise for weak magnetic signal detection. Third, researchers are actively considering arrays of coupled SQUIDs to boost the sensitivity of these bistable sys-tems even further.

With an eye to electromagnetic communications, a recent study³⁵ reported using SR to improve detection of a weak carrier signal, operated in both amplitude- and frequency-modulated modes (AM and FM). An interesting twist here is that the detec-tor was a bistable system (known as Chua's circuit), the two states of which were chaotic. Although there is no obvious advantage in using a chaotic detector, chaos does not impede the effect^{36,37}, illustrating yet again the generality of SR. Possibly indicative of the approaching technological exploitation, both Chua's circuit and the Quantum Magnetics SQUID have been implemented on silicon chips.

Physicists at the Georgia Institute of Technology (K.W. *et al.*, unpublished results) are studying the role that SR could play in enhancement of trigger-threshold detectors. In practice all detectors have a minimum threshold; inputs smaller than this are 'invisible'. The traditional approach to detector design is to make this threshold as small as possible, while isolating the sys-tem from noise as much as possible. However, at some point each of these objectives becomes either impractical or prohibitively expensive. An economical alternative may be, for a given thresh-

FIG. 3 Results of our experiment with crayfish mechanoreceptors (filled squares) compared to the electronic Fitzhugh–Nagumo simulation (diamonds) and the theory given by equation (1) (solid curve)²¹. The horizontal axis represents externally applied, gaussian noise: hydrodynamic noise in the case of the mechanoreceptor, and electronic noise in the case of the neuron model. The crayfish data do not decrease rapidly for small noise because of the residual internal noise of the neuron.

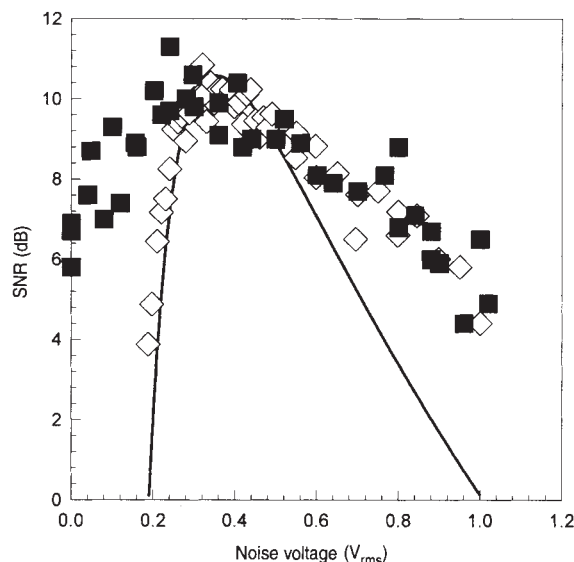
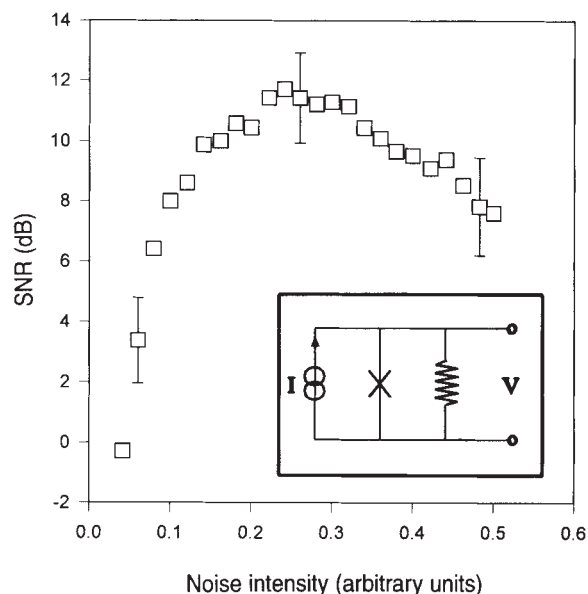


FIG. 4 Main figure, data from a numerical simulation of a trigger-reset system based on a Josephson junction as nonlinear element⁴⁸. Inset, circuit diagram of the Josephson system, consisting of an ideal junction (cross), quasiparticle resistance, and current source I which is the sum of three components—constant bias, weak periodic signal and noise. The output is the voltage, V .



old value, to allow enough noise to produce operation at the optimal value on the SNR curve. An example of SR in such a threshold device is shown in Fig. 4, generated from numerical simulations of a device that uses a superconducting Josephson junction as the nonlinear element.

Further examples

The possible role of SR is being pursued in other areas of science as well. One such study concerns human perception of simple ambiguous figures, for example the Necker cube³⁸, which may be viewed as a bistable process with an inherent noise. Bistability arises in the perception of which one of a pair of diagonally opposed corners of the transparent cube is in the foreground, and the perception spontaneously switches between the two alternatives. The switching is a random (or possibly a low-dimensional chaotic) process, and the fluctuation intensity can be controlled by the size and/or aspect ratio of the cube³⁹. A weak 'signal' can be introduced by moving a marker sinusoidally along a diagonal joining the two corners. In this way the eye, and hence the observer's attention, is weakly biased alternately between the two possible perceptions. Recently researchers demonstrated SR in a neural-network simulation of this process⁴⁰; an experiment by this

same group using human subjects is in progress.

A controversial issue, currently the focus of an intense public health debate, is the effect (if any) of extremely low frequency electromagnetic fields on living tissue^{41–43}. Theoretical estimates consistently predict the interaction energies of such fields after penetrating tissue to be up to three orders of magnitude smaller than the average energy of thermal fluctuations. How could such low-level coherence affect living cells⁴⁴? A concise review of this problem, including a summary of recent experimental results, is given in ref. 45. Recently, it was suggested that SR may play a role^{45–47}: voltage-sensitive ion channels in cell membranes behave like threshold devices in regard to external fields, randomly switching between open and closed states in response to thermal fluctuations. If SR is relevant, the effect of weak extremely low frequency electromagnetic fields might be greatly amplified. Whether any such enhancement is large enough to have significant biological ramifications is at this stage purely speculative. □

K. Wiesenfeld is at the School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. F. Moss is at the Department of Physics and Astronomy, University of Missouri at St Louis, St Louis, Missouri 63121, USA.

- Benzi, R., Sutera, S. & Vulpiani, A. *J. Phys.* **A14**, L453–L457 (1981).
- Nicolis, C. *Tellus* **34**, 1–9 (1982).
- Benzi, R., Parisi, G., Sutera, A. & Vulpiani, A. *Tellus* **34**, 10–18 (1982).
- Ghil, M. & Childress, S. *Topics in Geophysical Fluid Dynamics* part IV (Springer, Berlin, 1987).
- Winograd, I. J. *et al. Science* **258**, 255–260 (1992).
- Imbrie, J., Mix, A. C. & Martinson, D. G. *Nature* **363**, 531–533 (1993).
- Matteucci, G. *Clim. Dyn.* **6**, 67–81 (1991).
- Nicolis, C. *J. statist. Phys.* **70**, 3–14 (1993).
- Fauve, S. & Heslot, F. *Phys. Lett.* **97A**, 5–7 (1983).
- McNamara, B., Wiesenfeld, K. & Roy, R. *Phys. Rev. Lett.* **60**, 2626–2629 (1988).
- Simon, A. & Libchaber, A. *Phys. Rev. Lett.* **68**, 3375–3378 (1992).
- Moss, F., Bulsara, A. & Shlesinger, M. F. (eds) *J. statist. Phys.* **70**, 1–512 (1993).
- McNamara, B. & Wiesenfeld, K. *Phys. Rev.* **A39**, 4854–4869 (1989).
- Jung, P. & Hänggi, P. *Europhys. Lett.* **8**, 505–510 (1989).
- Gammaitoni, L., Marchesoni, F., Menichella-Saetta, E. & Santucci, S. *Phys. Rev. Lett.* **62**, 349–352 (1989).
- Jung, P. & Hänggi, P. *Phys. Rev.* **A44**, 8032–8042 (1991).
- Dykaman, M. I., Mannella, R., McClintock, P. V. E. & Stocks, N. G. *Phys. Rev. Lett.* **65**, 2606–2609 (1990).
- Jung, P. *Phys. Rep.* **234**, 175–295 (1993).
- Gammaitoni, L., Marchesoni, F., Menichella-Saetta, E. & Santucci, S. *Phys. Rev. Lett.* **71**, 3625–3628 (1993).
- Neiman, A. & Schimansky-Geier, L. *Phys. Rev. Lett.* **72**, 2988–2991 (1994).
- Wiesenfeld, K., Pierson, D., Pantazelou, E., Dames, C. & Moss, F. *Phys. Rev. Lett.* **72**, 2125–2129 (1994).
- Stocks, N. G., Stein, N. D. & McClintock, P. V. E. *J. Phys.* **A26**, L385–L390 (1993).
- Bulsara, A. R., Lowen, S. B. & Rees, C. D. *Phys. Rev.* **E49**, 4989–5000 (1994).
- Gingl, Z., Kiss, L. B. & Moss, F. *Europhys. Lett.* (in the press).
- Zhou, T., Moss, F. & Jung, P. *Phys. Rev.* **A42**, 3161–3169 (1990).
- Longtin, A., Bulsara, A. & Moss, F. *Phys. Rev. Lett.* **67**, 656–659 (1991).
- Longtin, A. *J. statist. Phys.* **70**, 309–327 (1993).
- Douglass, J. K., Wilkens, L., Pantazelou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
- Braun, H. A., Wissing, H., Schäfer, K. & Hirsch, M. C. *Nature* **367**, 270–273 (1994).
- Pantazelou, E., Dames, C., Moss, F., Douglass, J. & Wilkens, L. *Int. J. Bifurc. Chaos* (in the press).
- Moss, F., Douglass, J. K., Wilkens, L., Pierson, D. & Pantazelou, E. *Ann. N.Y. Acad. Sci.* **706**, 26–41 (1993).
- Hibbs, A., Jacobs, E. W., Bekkedahl, J., Bulsara, A. & Moss, F. in *Noise in Physical Systems and 1/f Fluctuations* (eds Handel, P. & Chung, A. L.) 720–723 (AIP Conf. Proc. Vol. 295, Am. Inst. Phys., New York, 1993).
- Hibbs, A. D. *et al. J. appl. Phys.* (in the press).
- Löfstedt, R. & Coppersmith, S. N. *Phys. Rev. Lett.* **72**, 1947–1950 (1994).
- Anishchenko, V. S., Safonova, M. A. & Chua, L. O. *Int. J. Bifurc. Chaos* **4**, 441–446 (1994).
- Anishchenko, V. S., Safonova, M. A. & Chua, L. O. *Int. J. Bifurc. Chaos* **2**, 397–401 (1992).
- Carroll, T. L. & Pecora, L. M. *Phys. Rev. Lett.* **70**, 576–579 (1993).
- Necker, L. A. *Phil. Mag.* **1**, 329–337 (1832).
- Kelso, J. A. S. *et al. in Multistability in Cognition* (eds Krouse, P. & Stadler, M.) (Springer, Berlin, in the press).
- Riani, M. & Simonotto, E. *Phys. Rev. Lett.* **72**, 3120–3123 (1994).
- Bennett, W. R. *Phys. Today* **47**, 23–29 (1994).
- Adair, R. K. *Phys. Rev.* **A43**, 1039–1048 (1991).
- Tenforde, T. S. A. *Rev. publ. Health* **13**, 173–196 (1992).
- Weaver, J. C. & Astumian, D. *Science* **247**, 459–462 (1990).
- Polk, C. in *Proc. 13th Nat. Council Radiat. Protection & Measmt* (National Council on Radiation Protection, Washington DC, in the press).
- Kruglikov, I. L. & Dertinger, H. *Bioelectromagnetics* (in the press).
- Makeyev, V. M. *Biophysics* **38**, 189–195 (1993).
- Wiesenfeld, K. & DeMino, K. in *Proc. 2nd Int. Workshop on Stochastic Resonance*, Nuovo Cim. (supplement) (ed. Mannella, R.) (in the press).

ACKNOWLEDGEMENT. We are grateful to M. F. Shlesinger for valuable discussions and continuing encouragement.

Acetylcholine receptor channel imaged in the open state

Nigel Unwin

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The structure of the open-channel form of the acetylcholine receptor has been determined from electron images of *Torpedo* ray postsynaptic membranes activated by brief (<5 ms) mixing with droplets containing acetylcholine. Comparison with the closed-channel form shows that acetylcholine initiates small rotations of the subunits in the extracellular domain, which trigger a change in configuration of α -helices lining the membrane-spanning pore. The open pore tapers towards the intracellular membrane face, where it is shaped by a 'barrel' of α -helices having a pronounced right-handed twist.

THE brain functions by sending electrical signals across different synapses connecting cells of the nervous system. Neurotransmitter-gated ion channels, of the acetylcholine (ACh) and glutamate receptor families, are the protein molecules in the postsynaptic membranes that control the rapid propagation of these signals between most cells. The fast-acting channels open transiently upon stimulation by the neurotransmitter and ions flow through them down their electrochemical gradients, leading to a change in membrane potential. How the opening is accomplished is not known, yet is fundamental to our understanding of synaptic communication.

The nicotinic ACh receptor, a 290K protein at the nerve-muscle synapse, is the most thoroughly studied neurotransmitter-gated ion channel¹⁻⁴. It consists of a ring of five similar subunits ($\alpha\beta\gamma\delta$), delineating through the membrane a central pathway for the ions. The pathway is narrow across the membrane, but widens into a ~ 20 Å diameter cylinder that extends ~ 65 Å into the synaptic cleft and ~ 15 Å into the interior of the cell⁵⁻⁷. The narrow, membrane-spanning portion, the pore, is lined by five α -helical segments, M2, one from each of the polypeptide chains^{2,4}. These segments in the absence of ACh appear to come together near the middle of the membrane and form the gate of the channel⁷. The gate opens upon binding of ACh to distant sites on the two α -subunits, and closes again when ACh is depleted from the synaptic cleft or when, in the continued presence of ACh, desensitization occurs. The muscle-type channels desensitize quickly (time constant: ~ 50 – 100 ms at 20 – 22 °C)^{8,9} but slowly compared to their opening rate (time constant: ~ 20 μ s)⁹⁻¹¹, and achieve close to a 100% probability of being open during initial activation with high concentrations of ACh⁸⁻¹¹.

Recently it has become possible to record images of ACh receptors, in which the open-channel form should predominate, by mimicking the initial transient activity at the synapse. The experiment (Fig. 1) involves brief mixing of spray droplets containing ACh with isolated postsynaptic membranes; rapid freezing to trap the structural response; and electron microscopy of the frozen membranes. I describe here the structure of the channel at 9 Å resolution, determined from the images after such activation and trapping, and compare it with the structure of the closed channel determined earlier at the same resolution⁷. The map reveals an altered configuration of M2 α -helical segments, encircling an open pore, and localized disturbances at the presumed level of the two ACh binding sites which are communicated to the membrane-spanning portion through small rotations of the subunits.

Millisecond mixing and freezing

ACh receptor-rich postsynaptic membranes were obtained from the muscle-derived electric organ of the *Torpedo* ray. The mem-

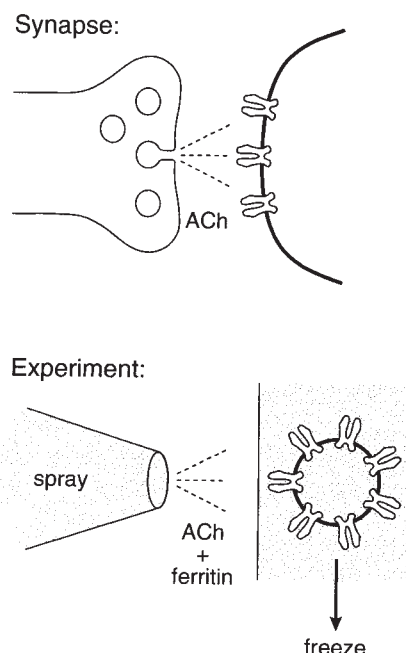
branes, after isolation and resuspension in a low-salt solution, crystallize in the form of long tubular vesicles¹². These tubes are composed of receptors, peripheral proteins and lipid molecules arranged on a p2 surface lattice in an outside-out orientation (Fig. 2a). A single image of a tube allows a complete three-dimensional structure to be determined, because the tube has overall helical symmetry¹³.

An atomizer spray coupled to a rapid freezing device was developed to achieve ACh-receptor reaction times of less than 5 ms at 5 °C (ref. 14), in order to trap the open channels and minimize desensitization. Solutions containing the tubes were applied to holey carbon support grids, which were blotted to produce thin aqueous films and plunged by free-fall into ethane cooled by liquid nitrogen. Spray droplets, containing 100 mM ACh and ferritin marker particles, were made to impinge on the grids 5 ms before they hit the ethane surface, thus allowing brief mixing with the film contents before the freezing. Tubes were photographed in areas of the frozen films near the edges of the spreading droplets, where the concentration of ferritin particles was typically 1/10–1/100th of the concentration in the spray. The concentration of ACh molecules in these areas would have risen by diffusion¹⁴ well beyond that needed to achieve a high probability of opening (~ 50 μ M)^{8,11}.

To check that the receptors in the presence of the ferritin had been affected by ACh, their structure was determined from each tube at only 25 Å resolution, and compared to their structure determined earlier from non-activated tubes⁷. Correlation coefficients were calculated between the two sets of densities in successive sections from a tube radius of 260 Å to 380 Å, corresponding to different levels along the receptor's length. With all such ACh-exposed tubes, the coefficients were highest when there were small, but reproducible, rotations between the two maps (Fig. 2b, filled circles). The rotations were smaller and of a different nature to those obtained from desensitized receptors in tubes exposed for several seconds to ACh (Fig. 2b, empty circles); no systematic effect could be detected from tubes with ferritin only present in the spray (data not shown). Therefore, the brief exposure to varying concentrations of ACh before freezing had induced a consistent and unique structural change, as would be expected if a high proportion of receptors had been trapped in the open-channel form.

Each of 28 tubes imaged in the vicinity of the spreading droplets yielded a set of Fourier terms extending to a resolution of 9 Å. These terms were averaged to improve the signal to noise ratio, as in the earlier structure determination from non-activated tubes⁷. In both structure determinations (Table 1), the averaging was done on >50,000 molecules, using tubes of the same helical family imaged under the same conditions (see legend to Fig. 2 for details). Comparable errors in the phases, over successive resolution zones (see legend to Fig. 2), demonstrated

FIG. 1 Channel-opening event at the synapse (above) and the experiment to capture this event *in vitro* (below). In the experiment, the spraying of ACh-containing droplets onto an aqueous film containing isolated postsynaptic membranes (tubular vesicles) mimics the release of neurotransmitter onto the target cell, and rapid freezing traps the channels in the open state. Ferritin marker particles, included in the spray solution, reveal the regions of the film where the droplets have impinged and spread, and so identify the membranes containing open channels.



that the activation did not introduce appreciable loss of crystallinity.

Channel mouth

A 9 Å map of the ACh-activated receptor was calculated from the averaged Fourier terms in sections normal to the axis of the channel, and then compared to the map of the non-activated receptor⁷. The two structures are most alike at the extracellular mouth of the channel: from the tips of the subunits to about a third of the way down towards the membrane (Fig. 3a, b). Within this region, the water-facing surfaces match fairly closely, and the two sets of densities are more similar (mean correlation coefficient: 0.91) than elsewhere in the receptor. Extensive subunit-subunit contacts are made at the channel mouth, which may limit the possibilities for rapid conformational change.

The two structures show greater differences beginning ~45 Å from the membrane, slightly above the region (A, Fig. 2b), where the ACh-binding sites are thought to be located^{7,15,16}, continuing along the shaft (S, Fig. 2b) connecting this level to the membrane, and around the pore itself. The disturbances, initiated at the ACh-binding sites and propagated to the membrane, appear as small net displacements around the channel axis at low resolution, which correlate with the negative-positive rotations indicated in Fig. 2b.

ACh-binding sites

The ACh-binding sites in the α -subunits are not equivalent, because they are shaped partly by the neighbouring γ - or δ -subunits¹⁷⁻²⁰ and have different affinities for ACh^{21,22}. In the non-activated structure at the presumed level of these sites, the α -subunits were seen to contain sets of three rods of density, typical of α -helices, running roughly normal to the membrane plane⁷. The other subunits shared these features. However, unique to both α -subunits at the centres of the set of rods were equal-sized cavities, which may be the actual ACh-binding pockets⁷. In the activated structure, the α -subunit next to the δ -subunit, α_δ , is noticeably altered because the cavity it had has disappeared; otherwise, corresponding subunits have similar individual features in the two maps.

Figure 3c is a cross-section through the activated structure at

the presumed level of the binding sites, with peaks in the contours indicating the positions of the rods, and crosses marking the former positions of the two cavities, only one of which (in α_γ) is still visible. At α_δ , where the cavity has disappeared, the encircling peaks are displaced anticlockwise from their former positions (blue spots). If the peaks are traced at successive levels to give a three-dimensional picture (Fig. 3d), and compared to the former traces (Fig. 3d, inset), it is seen that these displacements originate from a concerted twisting of the rods. The largest changes are above the (former) cavity, where each rod has rotated anticlockwise by up to ~28°. Smaller changes in the positions of the rods, beneath the cavity, appear to initiate a small clockwise rotation of the subunit that continues down to the membrane (see below).

The localized twisting in α_δ displaces the rods circumferentially, and so must affect the neighbouring subunits. Indeed, it can be shown that this disturbance is part of a coordinated change in the vicinity of the binding sites by comparing the two structures in sectors of 72° (rather than 360°, as before), centred over successive subunits. The sector comparison indicates that the disturbance at α_δ contributes the most to the varying rotational disparities between sections around this level in Fig. 2b, and that the effects diminish going anticlockwise around the ring (Fig. 2b, box). In contrast, the β -subunit (between the

TABLE 1 Crystallographic data

	-ACh	+ACh (<5 ms)
Number of tubes*	26	28
Receptors	55,000	51,100
Resolution (Å)	8.7	8.7
Defocus range (Å)	7,000-18,800	6,200-16,400
Layer-lines	230	234
Fourier terms	9,126	9,022
2-Fold phase error (°)†	10.6	12.4

* All tubes have (-16, 6) helical symmetry^{6,7}; the p2 surface lattice has dimensions: $a=80.2$ Å, $b=145.5$ Å, $\gamma=114.0^\circ$ (at tube radius: 290 Å).

† Amplitude-weighted phase error against 0° or 180°; see also legend to Fig. 2.

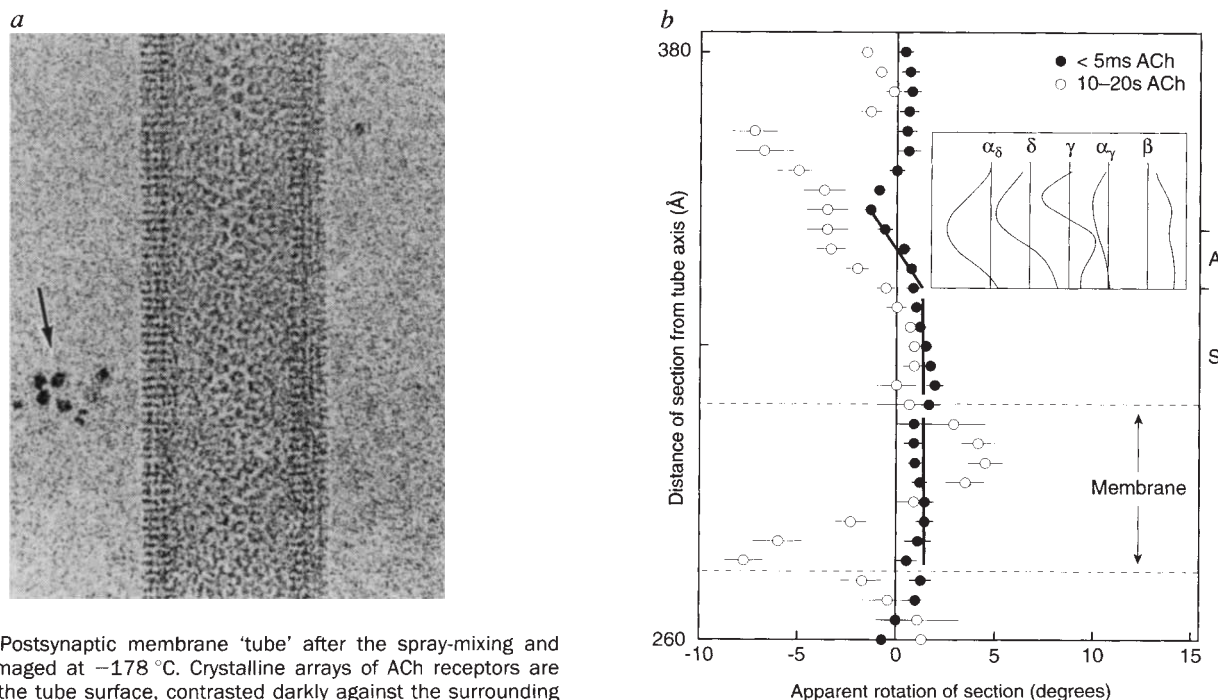


FIG. 2 *a*, Postsynaptic membrane 'tube' after the spray-mixing and freezing, imaged at -178°C . Crystalline arrays of ACh receptors are visible on the tube surface, contrasted darkly against the surrounding lipid and ice. The extracellular domains of the receptors face outwards, giving the edge of the tube a striated appearance. The presence of ferritin particles (arrow) in the vicinity of the tube indicates that the area is near the edge of a spreading droplet, and that the receptors have therefore been exposed to ACh. The tube diameter is 760 Å. *b*, Structural changes in the ACh receptor after spray-mixing and freezing (filled circles), and prolonged application of ACh to induce desensitization (unfilled circles), detected at 25 Å resolution in three-dimensional maps from individual tubes. Each map was compared, section by section at 4 Å intervals, with the map of the receptor determined from non-activated tubes⁷. Shown are the rotations required to give maximum correlations between sections at successive levels along the receptor's length (extracellular end uppermost), measured from the central axis of the tube. The data were obtained from 37 maps (sprayed tubes) and 8 maps (tubes to which 5 mM ACh had been added 10–20 s before freezing); the bars are standard errors of the mean. A denotes the level of the ACh binding sites and S the level of the shaft connecting this level to the membrane. Lines drawn through the datapoints emphasize distinct disturbances, initiated by ACh, at the levels of the ACh-binding sites, the shaft and the membrane. The box compares the results obtained by correlating the densities in sectors of 72° (also at 25 Å resolution) to show how the net rotational differences at level A (and slightly above) are distributed among the individual subunits, going anticlockwise around the ring; the scales for each subunit, either side of the vertical lines, are as in the main figure.

METHODS. Tubes were suspended in 100 mM sodium cacodylate, pH 6.8, and sprayed with 100 mM acetylcholine chloride and 2 mg ml⁻¹ ferritin at the same pH. The tube suspension was applied in 5 µl aliquots to glow-discharged holey carbon films, supported on rigid 300 mesh grids. Rapid freezing was accomplished by a guillotine-type plunger, situated in a cooled high humidity environment, using liquid ethane (at $\sim -160^\circ\text{C}$) as the coolant. Excess solution was blotted off the grid from the side opposite that to which it was applied, and then the grid was plunged by free-fall into the liquid ethane. The plunger incorporated an atomizer spray¹⁴ to deliver a pulse of $\sim 1\text{ }\mu\text{m}$ diameter droplets onto the falling grid. The spray pulse was activated by a photodiode sensor, which detected the displacement of the plunger shaft. The release point of the grid was 35 mm above the nozzle, which was 4 mm above the surface of the ethane, giving 5 ms maximum reaction time between the ACh and the receptors. The temperature of the reaction was 5°C , measured by a thermocouple attached to a falling grid. Grids were

stored in liquid nitrogen and examined within 1.5 years of freezing, using a Gatan cryoholder and a Philips EM420 electron microscope equipped with a low dose kit and operating at 120 kV. Images were recorded at X36,000, of tubes lying within holes in the carbon support film. The electron dose for each image was ~ 6 electrons per Å², and a wide range of underfocus was used to ensure that all spacings, at 7% amplitude contrast⁴⁶, were well sampled⁴⁷. Tubes belonging to the $(-16, 6)$ helical family^{6,7}, and surrounded by ferritin particles, were selected visually and by the appearance of the layer-lines in their optical diffraction patterns. The presence of ferritin particles was an important requirement, in addition to the criteria discussed earlier⁷: no effect of ACh (Fig. 2b) could be detected in regions of tubes located further than $\sim 1\text{ }\mu\text{m}$ from the nearest ferritin particle. Selected tubes were densitometered and analysed further by computer, essentially as in ref. 7. Briefly, the Fourier terms were obtained along the layer-lines extracted from stretches of tubes equal to integral multiples of the repeat distance; layer-lines from the different stretches were then merged to the same centrosymmetrical origin, taking account of the different contrast transfer functions and the qualities of the data along the Z (tube axis) and R (radial) directions from each stretch. The structure was calculated directly, without rescaling the amplitudes to compensate for the fading at higher resolution, by Fourier-Bessel inversion and Fourier synthesis. 5-fold averaging was done to obtain the best estimate of the structure common to all five subunits in the membrane-spanning region; as before⁷, only one pore-lining rod could be resolved in each subunit before the averaging; no significant non-5-fold feature could be detected in this portion of the receptor. The final dataset was derived from 37 stretches along 28 sprayed tubes. A plot of the sum of the moduli of the contrast transfer functions, from the 28 images, was almost identical to the plot obtained from the non-activated tubes⁷. The overall amplitude-weighted phase error: $\sum |F| \delta\phi / \sum |F|$ (where $\delta\phi$ is the difference between the observed phase and the symmetry-imposed phase of 0° or 180°) was 12.4° . The errors obtained from layer-lines assessed in annular zones of increasing resolution were: 3.5° (68.0–34.5 Å); 8.6° (24.5–22.8 Å); 15.4° (22.8–17.0 Å); 29.1° (17.0–13.7 Å); 31.1° (13.7–11.4 Å); 31.8° (11.4–9.8 Å); 31.9° (9.8–8.7 Å); the corresponding figures for the non-activated structure were: 2.9° ; 6.2° ; 13.9° ; 23.0° ; 24.5° ; 30.4° ; 35.1° , with roughly the same number of measurements (>120) made in each zone.

α -subunits; but see legend to Fig. 3) contributes nothing to the varying rotational disparities, and is displaced in the opposite sense to α_δ . The β -subunit is therefore the least flexible in the vicinity of the binding sites and, instead of deforming, moves away from α_δ . This movement, in Fig. 3c, has led to small shifts

in the peaks in density away from their former locations, following the direction of the arrow.

ACh induces a similar disturbance in α_γ as in α_δ (Fig. 2b, box), but affects this subunit more weakly, presumably because α_γ is tightly coupled to the β -subunit. On the basis of the

details just described, it seems that the tight coupling and imposed rigidity have together enabled the disturbance initiated at α_γ to draw the β -subunit away from α_δ (arrow, Fig. 3c), opening up the structure to allow the twisting. This combined action coordinates the structural responses of the two

α -subunits, and so is likely to be central to the cooperative mechanism of channel opening²². However, the ACh molecule itself and other fine features have not been resolved, and Fig. 2b indicates that the interactions at this level are quite complex.

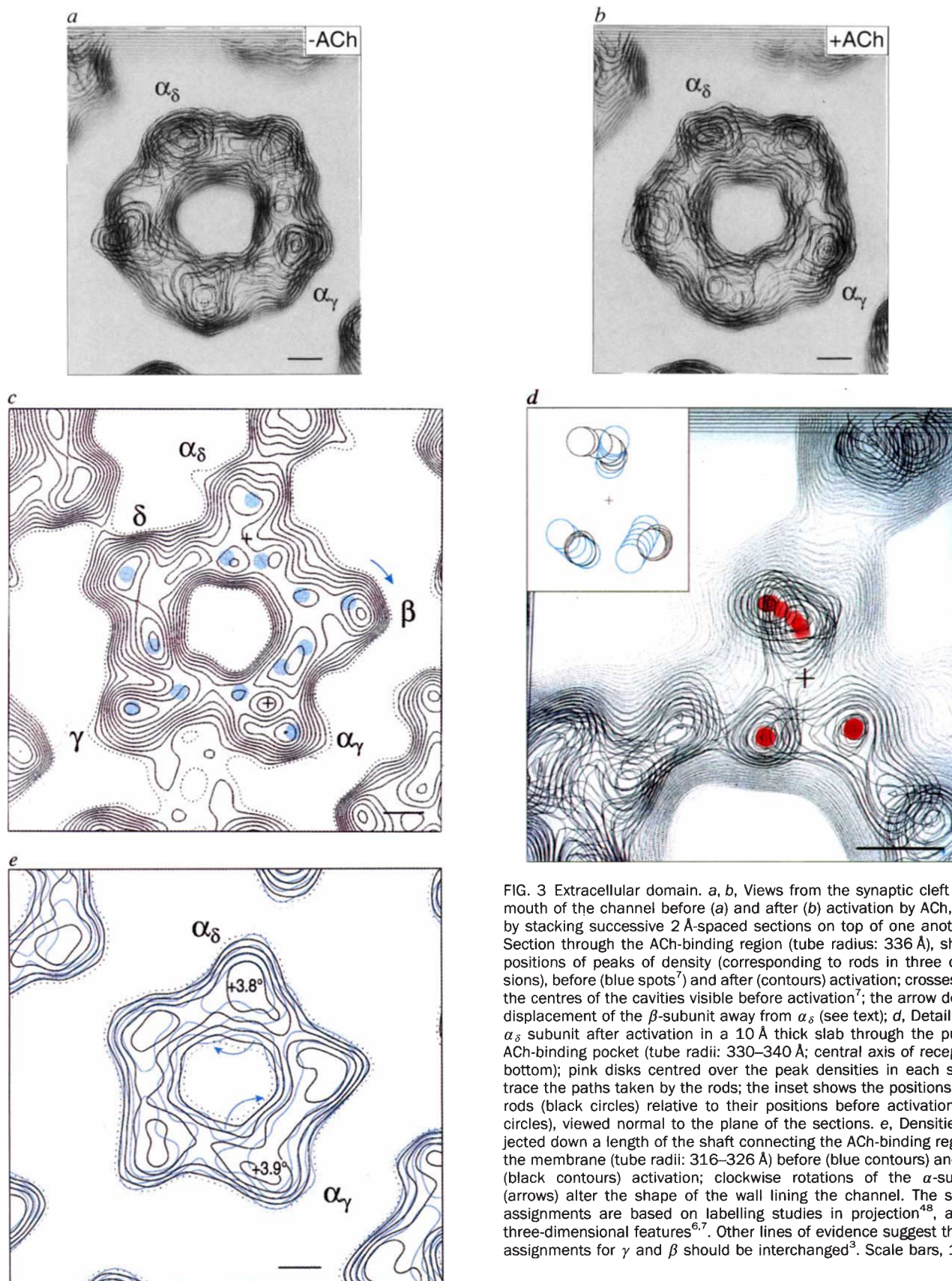


FIG. 3 Extracellular domain. a, b, Views from the synaptic cleft of the mouth of the channel before (a) and after (b) activation by ACh, made by stacking successive 2 Å-spaced sections on top of one another; c, Section through the ACh-binding region (tube radius: 336 Å), showing positions of peaks of density (corresponding to rods in three dimensions), before (blue spots) and after (contours) activation; crosses mark the centres of the cavities visible before activation⁷; the arrow denotes displacement of the β -subunit away from α_δ (see text); d, Detail of the α_δ subunit after activation in a 10 Å thick slab through the putative ACh-binding pocket (tube radii: 330–340 Å; central axis of receptor at bottom); pink disks centred over the peak densities in each section trace the paths taken by the rods; the inset shows the positions of the rods (black circles) relative to their positions before activation (blue circles), viewed normal to the plane of the sections. e, Densities projected down a length of the shaft connecting the ACh-binding region to the membrane (tube radii: 316–326 Å) before (blue contours) and after (black contours) activation; clockwise rotations of the α -subunits (arrows) alter the shape of the wall lining the channel. The subunit assignments are based on labelling studies in projection⁴⁸, and on three-dimensional features^{6,7}. Other lines of evidence suggest that the assignments for γ and β should be interchanged³. Scale bars, 10 Å.

Shaft

The shaft is a roughly 25 Å-long region of the receptor, connecting the binding-site region to the membrane-spanning portion in which the subunits appear as distinct columns of density aligned parallel to the axis of the channel⁶. Figure 3e compares the activated and non-activated (blue) shaft structures projected down the axis. Small clockwise rotations of the α -subunits (arrows), that appear to be initiated at the binding sites, are propagated down the shaft. These show in Fig. 3e as changes in the distributions of the densities composing both α -subunits, and as distortions to the shape of the inner channel wall (which in the non-activated structure is pentagonally symmetrical).

No distinctive secondary structure is resolved in the shaft region of the receptor, so the changes in this region are not as obvious as they are in the vicinity of the binding sites. By isolating the densities composing each subunit in 25 Å diameter cylinders around its centre of mass, and comparing these densities in successive sections down the shaft (beginning 8 Å below the level in Fig. 3c), it is found that the α -subunits have to be rotated by $3.8^\circ \pm 0.8$ s.d. (α_δ) and $3.9^\circ \pm 1.1$ s.d. (α_γ) to achieve highest correlations between the two maps (Fig. 3e). The same procedure, applied to the other subunits, indicates smaller rotations ($<2^\circ$), which are, however, also in a clockwise orientation. Thus, the two α -subunits, to which ACh binds to trigger the initial disturbances, also play the

primary role in transmitting the effects of these disturbances to the membrane.

Membrane-spanning structure

The membrane-spanning portion of the activated receptor reveals five rods of density, 25–35 Å long, forming a wall around a pore, and a star-shaped outer rim of continuous density, which faces the lipids. The rods and the rim are arranged symmetrically around the pore, consistent with the high sequence homology among the putative transmembrane segments²³. Figure 4a shows these features, after imposing 5-fold symmetry to obtain the best estimate of the structure common to the subunits. The rods can be identified with the α -helical stretches of amino acids, M2, because they are exposed to the lumen of the pore throughout their length^{2,4}. The rim of continuous density seems most likely to be composed of β -sheet according to its appearance⁷; other recent measurements have interpreted the remaining transmembrane segments to be mainly β -structure and turns²⁴ or mainly α -helical²⁵.

The M2 helices bend or 'kink' about their mid-points, and so are arranged differently in the two leaflets of the lipid bilayer. In both leaflets they slope radially (by $\sim 12^\circ$) towards the axis of the pore, but in the lower (intracellular) leaflet they are predominantly at a tangent to it (Fig. 4b, upper left). Neighbouring helical segments in the upper leaflet also are sufficiently apart

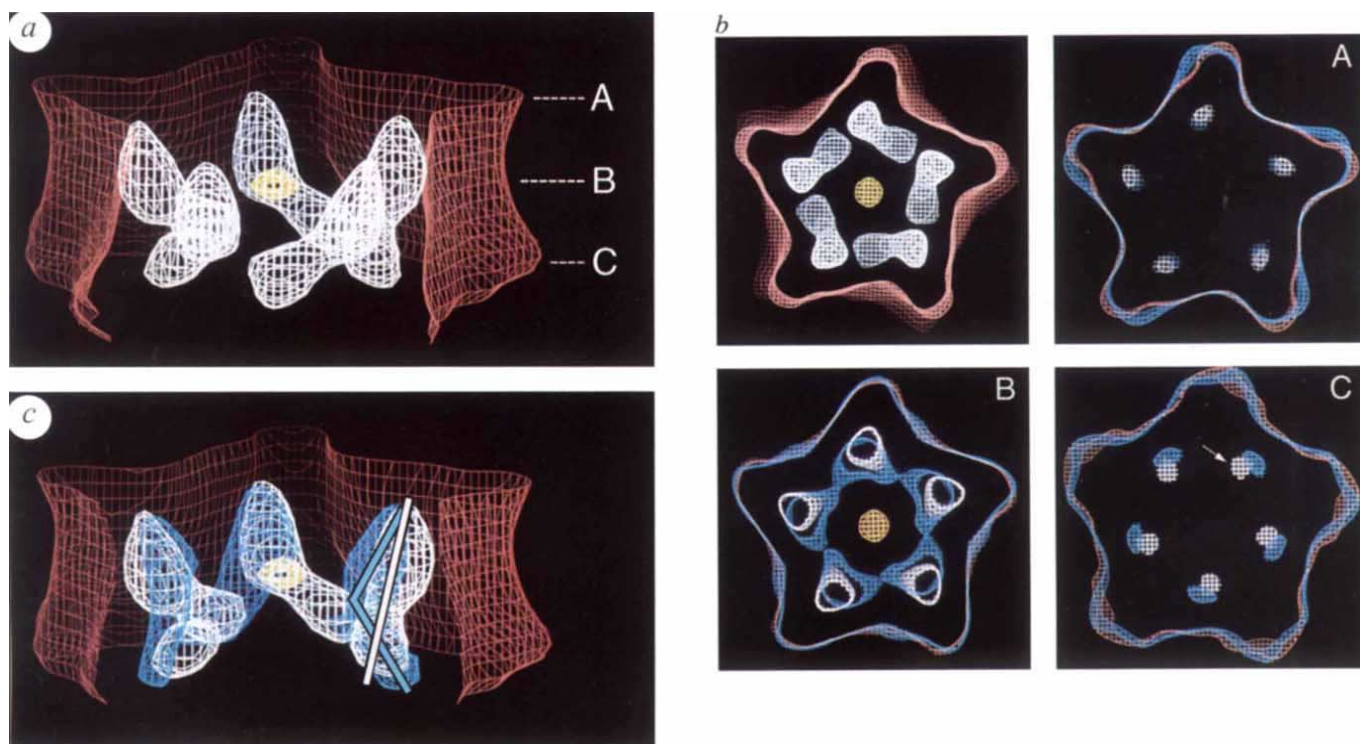


FIG. 4 Membrane-spanning structure. a, Open pore showing the M2 helices (white) and the lipid-facing 'rim' (brown); the helical segments bend abruptly near the middle of the membrane and twist around the central axis in the lower (intracellular) leaflet; b, Axial view of open pore (upper left) and sections at the levels A, B and C in a, comparing details with those of the closed pore (blue); the helical segments of the open pore are tangential to the axis in the lower leaflet (arrow in C), and so have moved the most in this region; c, Open pore with the helices of the closed pore (blue) superimposed: the helices around the open pore (white line) are oriented so that they constrict the ion path most at the lower membrane face, whereas the helices around the closed pore (blue line) are oriented so that they point inwards near the middle of

the membrane to make the closed gate⁷. The front face of the rim has been removed in a and c to reveal the helices more clearly. A high-density feature (yellow) on the axis near the middle of the membrane may reflect accumulation of ACh (which acts as a channel blocker at millimolar concentrations⁴⁹) and the proximity at this level of bulky, strongly scattering side-chains. The membrane-spanning portion shown (tube radii: 276–300 Å) corresponds to the hydrophobic core, lying between the phospholipid headgroups⁷. The map grid units correspond to 1 Å (horizontal) and 2 Å (vertical), and the densities are displayed at 50% (rim) and 90% (helices and central feature) of the maximum density.

(up to 21 Å centre-to-centre separation) to allow exposure to the pore of side-chains composing the outer rim, whereas in the lower leaflet they come too close (~10.5 Å centre-to-centre separation) for this to happen, and must be touching one another.

The M2 segments in the intracellular leaflet slope at an angle of ~45° tangential to the axis of the pore. Together they form a 'barrel' of α -helices having a pronounced right-handed twist, that resembles the right-handed barrels of pore-lining α -helices found in the bacterial toxin B-pentamers^{26,27}, and in heavy riboflavin synthase²⁸. However, in the case of the ion channel, it seems likely that the motif is stabilized mainly by contacts around the barrel, with minimal interactions involving residues from the rim. The lower portions of the helices are further from the rim than they are in the closed configuration (see below), and the lower helix residues facing the rim (on the opposite face to those in Fig. 5c) are polar. The rigidity of a barrel, combined with a low stability, may be important in ensuring both precise permeation and fast gating kinetics.

Alternative configurations of M2

The structure of the 'open' pore is quite different from that of the closed pore⁷, where the helices are also kinked about their mid-points, but come closest to the central axis at the kinks (forming the gate) and slope roughly radially away from the axis above and below that point. Figure 4b compares the two structures in sections level with the middle of the membrane and with the two faces (closed pore in blue). The helices making the open pore are further from the axis at the middle of the membrane (B, Fig. 4b) and closer to the axis at the lower face (C, Fig. 4b). The greatest changes are at the lower face, where the helical segments move across the apex of the star formed by the outer rim to achieve the open pore arrangement (arrow in C, Fig. 4b). The rim has the same shape in both structures, suggesting it may act as a framework to partition the moving helices away from the lipids.

These changes in configuration of the helices are linked to a change in orientation of the kinks (Fig. 4c). Instead of pointing inwards (closed configuration), they have rotated over to the side, opening up the pore in the region where the gate was located. The rotations are in the same sense as those in the shaft. The kink therefore seems to act as a region of flexure between

two α -helical segments, accommodating an extended conformational change. Kinks between α -helical segments have been found to have analogous roles in other proteins²⁹. Clearly, the localized flexure provides a simple and effective way of altering the shape, size and environment of the pore.

The ability of the helices to move inside the outer rim seems to be helped by the fact they have two modes of association, neither of which is particularly extensive. The open pore association involves side-to-side interactions around the barrel, whereas the closed pore association involves interactions between exposed side-chains forming a tight central ring⁷. Apparently, these alternative associations reflect equally symmetrical arrangements of the helices around the pore, because the rods of density, after imposing 5-fold symmetry, are resolved with equal clarity in the two maps. There is no indication, in either structure, that the M2 segments contain a stretch of extended polypeptide chain of seven residues³⁰, but a shorter stretch³¹ would be compatible with the abrupt change of angle at the kink.

Ion permeation pathway

The open pore narrows fairly uniformly from the upper to the lower face of the membrane because of a constant radial inclination of the helices of ~12° (as indicated by the line drawn through the helix in Fig. 4c). The axes of the helices are about 18 Å from the pore axis at the upper face and 11.5 Å at the lower face, where the pore is most constricted. Assuming the figure of 11.5 Å for the distance of closest approach of the helix to the pore axis, and that hydroxyl-containing residues are projecting directly inwards at this point (see below), the minimum pore diameter would be 9–10 Å. This diameter is only slightly larger than the minimum pore diameter of mouse ACh receptors (8.4 Å)³², estimated by measurement of their permeability to organic cations of various size. The short extent of the constriction agrees with an estimate of 3–6 Å, based on reversal potential measurements from ACh receptors exposed to osmotic pressure gradients³³.

The uniform narrowing of the open pore towards the lower, N-terminal end of M2 agrees with the results of several types of experiment. Channel blocking reagents, such as chlorpromazine^{34,35}, triphenylmethylphosphonium³⁶ and 3-trifluoromethyl-3-*m*-iodophenyldiazirine³⁷ were found to

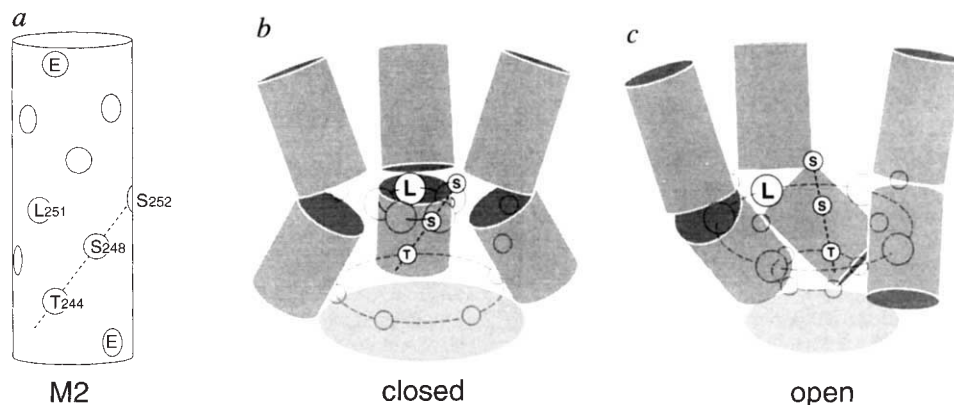


FIG. 5 Pore-lining segment, M2. *a*, Identified pore-facing amino-acid residues^{34–41} on an ideal α -helix (6.5 Å radius) viewed from inside the pore; the assignments, E, L251, S252, S248, T244 and E are for the *Torpedo* α -subunit; the glutamic acid residues, E, form part of the 'outer' (upper) and 'intermediate' (lower) rings of negative charge³⁸; polar pore-facing residues (serines and threonine) in the lower half of the helix lie

at an oblique angle (broken line) to the axis; all subunits have a leucine residue and polar residues at equivalent positions in the sequence. *b*, Closed⁷ and *c*, open configurations of M2, with the leucines and polar residues superimposed (see text); relative to *b*, the configuration in *c* has moved back the leucines, reoriented the line of polar residues, and brought T244 (and equivalent residues) closer to the axis of the pore.

penetrate to, and photolabel, residues located in the middle portion of M2, including a highly conserved leucine residue, L251 (refs 35, 37). Mutations reducing the net charge of residues composing the ring of negative charge at the lower end of M2 (E; Fig. 5a) were found to affect strongly ion flow, as if this ring is close to the narrowest portion of the pore³⁸. An analysis of the influence of mutations on the binding of the open-channel blocker, QX222, indicated that the pore tapers towards the lower end of M2 and drew attention to the likely importance of a line of polar pore-facing residues (S252, S248, T244; Fig. 5a)³⁹. Mutation of the threonine residue at the bottom of this line (T244) was found either to restrict or enhance ion flow depending on the volume of substituted residue, indicating that this threonine is at, or near to, the point of maximum constriction^{40,41}.

Alignment of M2 with sequence

The polar pore-facing residues in the lower portion of the ideal α -helix in Fig. 5a form a line that slopes by $\sim 40^\circ$ to its axis at the radius of their hydroxyl groups (6–7 Å). The inclination is similar in magnitude, but opposite in sense, to the slope of the M2 helices in this region of the open pore (Fig. 4a). Therefore the line of polar residues should lie almost parallel to the axis of the open pore, an orientation that would aid ion flow by giving optimal exposure of the hydroxyl groups⁴². Furthermore, threonine T244 must be close to the point where the open pore is most constricted^{40,41}. Identification of T244 with this level in the structure supports the tentative alignment of the three-dimensional densities with the amino-acid sequence⁷, suggesting that L251 is the gate-forming residue projecting from the kink: the kink and the constriction are about 10 Å apart (Fig. 4c), and a similar α -helical separation is predicted from the relative positions of these residues in the sequence.

Figure 5b, c shows the pore-facing residues in alignment with the two configurations of M2, according to the levels indicated for L251 and T244. It had been proposed⁷ that the helices, by bending towards the central axis, would allow the large leucine side chains to project inwards and associate in a tight ring, preventing the ions from crossing the membrane (closed pore, Fig. 5b); details of the activated receptor now imply that the helices, by bending tangentially to the central axis and associating side-to-side, would allow the polar surfaces to become more fully exposed, while moving the large side-chains away from the ion path (open pore; Fig. 5c).

Mechanism of opening

How does ACh weaken the 'leucine-ring' mode of association of the helices (Fig. 5b), in favour of their side-to-side association around the barrel (Fig. 5c), and open the channel? Although important details have not been resolved, the results of this study suggest some essential elements of the conformational change. First, ACh triggers distinct, localized disturbances at the binding sites in the two α -subunits, which are coordinated allosterically by the single subunit between them. Second, the effects of these disturbances are communicated, through small rotations of the subunits, to the structure in the membrane. Third, the M2 helices in the membrane transmit the rotations to the gate-forming side-chains, drawing them away from the central axis. The ring mode of association is thereby disfavoured, and the helices switch to the alternative side-to-side mode, creating an open pore.

This mechanism would account for the different photolabeling patterns obtained in the presence and absence of agonist³⁷, suggesting that closely packed hydrophobic residues near the middle of M2 may be 'splayed apart' upon opening. A feature of the channel, once opened, is that it is (weakly) stabilized across the narrow membrane-spanning portion by interactions between the helices around the barrel; so the exact nature of the rotations induced by the ligand are unlikely to be critical as long as they are sufficient to destabilize the gate. These properties may be important in interpreting the strong effects of some mutations involving residues within or (spatially) near to M2 in the ACh receptor⁴³ and in the homo-oligomeric α -7 nicotinic receptor⁴⁴, and the distinct ligand binding and channel specificities of chimeric receptors having swapped extracellular domains⁴⁵.

Conclusion

The experiment reported here, to capture and image the ACh receptor in the open-channel form, has given direct insight into how the receptor works. The confinement of major disturbances to the main functional regions, the development of small rotational motions to disrupt the gate, and the transient formation of an α -helical barrel around the open pore are new principles revealed. Together, they provide a structural basis for explaining the remarkably rapid and efficient response of this synaptic protein to the neurotransmitter, its precise permeation and its almost total impermeability to ions when the neurotransmitter has left^{1,8–11,21,32}. It seems likely that other neurotransmitter-gated channels would use similar principles to open their pores and let the ions through. □

Received 12 October; accepted 8 December 1994.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1992).
- Lester, H. A. A. *Rev. Biophys. biomolec. Struct.* **21**, 267–292 (1992).
- Karlin, A. *Curr. Opin. Neurobiol.* **3**, 299–309 (1993).
- Changeux, J.-P., Galzi, J.-L., Devillers-Thiery, A. & Bertrand, D. *Q. Rev. Biophys.* **25**, 395–432 (1992).
- Toyoshima, C. & Unwin, N. *Nature* **336**, 247–250 (1988).
- Toyoshima, C. & Unwin, N. *J. Cell Biol.* **111**, 2623–2635 (1990).
- Unwin, N. *J. molec. Biol.* **229**, 1101–1124 (1993).
- Dilger, J. P. & Brett, R. S. *Biophys. J.* **57**, 723–731 (1990).
- Matsubara, N., Billington, A. P. & Hess, G. P. *Biochemistry* **31**, 5507–5514 (1992).
- Colquhoun, D. & Sakmann, B. *J. Physiol. Lond.* **369**, 501–557 (1985).
- Colquhoun, D. & Ogden, D. C. *J. Physiol. Lond.* **395**, 131–159 (1988).
- Brisson, A. & Unwin, P. N. T. *J. Cell Biol.* **99**, 1202–1211 (1984).
- DeRosier, D. J. & Klug, A. *Nature* **217**, 130–134 (1968).
- Berriman, J. A. & Unwin, N. *Ultramicroscopy* **56**, 241–252 (1994).
- Herz, J. M., Johnson, D. A. & Taylor, P. J. *biol. Chem.* **264**, 12439–12448 (1989).
- Valenzuela, C. F., Weigen, P., Yguerabide, J. & Johnson, D. A. *Biophys. J.* **66**, 674–682 (1994).
- Blount, P. & Merlie, J. P. *Neuron* **3**, 349–357 (1989).
- Pedersen, S. E. & Cohen, J. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2785–2789 (1990).
- Sine, S. M. & Claudio, T. J. *biol. Chem.* **266**, 19369–19377 (1991).
- Czajkowski, C. & Karlin, A. J. *biol. Chem.* **266**, 22603–22612 (1991).
- Jackson, M. B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2219–2203 (1989).
- Sine, S. M., Claudio, T. & Sigworth, F. S. *J. gen. Physiol.* **96**, 395–437 (1990).
- Popot, J.-L. & Changeux, J.-P. *Physiol. Rev.* **64**, 1162–1239 (1984).
- Görne-Tschelnokov, U. et al. *EMBO J.* **13**, 338–341 (1994).
- Blanton, M. P. & Cohen, J. B. *Biochemistry* **33**, 2859–2872 (1994).
- Stein, P. E. et al. *Structure* **2**, 45–57 (1994).

- Sixma, T. K. et al. *Nature* **351**, 371–377 (1991).
- Ladenstein, R. et al. *J. molec. Biol.* **203**, 1045–1070 (1988).
- Gerstein, M., Schulz, G. & Chothia, C. *J. molec. Biol.* **229**, 494–501 (1993).
- Akabas, M. H., Stauffer, D. A., Xu, M. & Karlin, A. *Science* **258**, 307–310 (1992).
- Akabas, M. H., Kaufmann, C., Archdeacon, P. & Karlin, A. *Neuron* **13**, 919–927 (1994).
- Cohen, B. N., Labarca, C., Davidson, N. & Lester, H. A. *J. gen. Physiol.* **100**, 373–400 (1992).
- Dani, J. A. *J. Neurosci.* **9**, 884–892 (1989).
- Giraudat, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2719–2723 (1986).
- Revah, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4675–4679 (1990).
- Hucho, F. L., Oberthur, W. & Lottspeich, F. *FEBS Lett.* **205**, 137–142 (1986).
- White, B. H. & Cohen, J. B. *J. biol. Chem.* **267**, 15770–15783 (1992).
- Imoto, K. et al. *Nature* **335**, 645–648 (1988).
- Charnet, P. et al. *Neuron* **2**, 497–513 (1990).
- Imoto, K. et al. *FEBS Lett.* **289**, 193–200 (1991).
- Villarroel, A., Herlitz, S., Koenen, M. & Sakmann, B. *Proc. R. Soc. B243*, 69–74 (1991).
- Unwin, N. *Neuron* **3**, 665–676 (1989).
- Lee, Y.-H. et al. *Biophys. J.* **66**, 646–653 (1994).
- Galzi, J.-L. et al. *Nature* **359**, 500–505 (1992).
- Eiselé, J.-L. et al. *Nature* **366**, 479–483 (1993).
- Toyoshima, C. & Unwin, N. *Ultramicroscopy* **25**, 279–292 (1988).
- Unwin, P. N. T. & Henderson, R. *J. molec. Biol.* **94**, 425–440 (1975).
- Kubalek, E., Ralston, S., Lindstrom, J. & Unwin, N. *J. Cell Biol.* **105**, 9–18 (1987).
- Sine, S. M. & Steinbach, J. H. *Biophys. J.* **46**, 277–284 (1982).

ACKNOWLEDGEMENTS. I thank A. Auerbach, F. Sachs, G. Hess and my colleagues at the MRC Laboratory and the Scripps Research Institute, La Jolla for helpful discussions and encouragement. The Marine Station, Roscoff, France, kindly supplied the *T. marmorata* for this work. The research was supported, in part, by grants from the NIH and the CEE.

Discovery of shell-like radio-structure in SN1993J

J. M. Marcalde^{*†}, A. Alberdi[‡], E. Ros^{*},
P. Diamond[§], B. Schmidt[†], I. I. Shapiro[†],
L. Baath^{||}, R. J. Davis[¶], A. G. de Bruyn[#],
P. Elósegui[†], J. C. Guirado[‡], D. L. Jones[☆],
T. P. Krichbaum^{**}, F. Mantovani^{††}, R. A. Preston[☆],
M. I. Ratner[†], A. Rius^{‡‡}, A. E. E. Rogers^{§§},
R. T. Schilizzi^{|||}, C. Triglilio^{¶¶}, A. R. Whitney^{§§},
A. Witzel^{**} & A. Zensus[§]

^{*} Departamento de Astronomía, Universitat de València, 46100 Burjassot, Valencia, Spain

[†] Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

[‡] Instituto de Astrofísica de Andalucía, CSIC, 18080 Granada, Spain

[§] National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA

^{||} Onsala Space Observatory, 43992 Onsala, Sweden

[¶] Nuffield Radio Astronomy Laboratories, Macclesfield, Cheshire SK11 9DL, UK

[#] Netherlands Foundation for Research in Astronomy, 7990 AA Dwingeloo, The Netherlands

[☆] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

^{**} Max-Planck-Institut für Radioastronomie, Bonn 53010, Germany

^{††} Istituto di Radioastronomia, CNR, Bologna 40129, Italy

^{‡‡} Instituto de Astronomía y Geodesia, Facultad de Ciencias Matemáticas, 28040 Madrid, Spain

^{§§} MIT-Haystack Observatory, Westford, Massachusetts 01886, USA

^{|||} Joint Institute for VLBI in Europe, 7790 AA Dwingeloo, The Netherlands

^{¶¶} Istituto di Radioastronomia, CNR, Noto, Italy

SUPERNOVA explosions are poorly understood, partly because of difficulties in modelling them theoretically¹, and partly because there have been no supernovae observed in our Galaxy since the invention of the telescope. But the recent discovery² of supernova SN1993J in the nearby galaxy M81 offers an opportunity to investigate the evolution of the remnant, and its interaction with the surrounding interstellar medium, at high resolution. Here we present radio observations of SN1993J, made using very-long-baseline interferometry, which show the development of a shell structure. This 8-month-old radio shell is the youngest ever discovered in a supernova. The data suggest that the supernova explosion and the expanding shell of the remnant have nearly spherical symmetry, with small deviations where some parts of the shell are brighter than others. If these deviations arise because of variations in the density of the shell, this may reconcile earlier reports of symmetric radio emission³ with the observed optical asymmetry^{4,5}, as the density variations could easily cause the latter. We infer that the radio emission is generated at the interface⁶⁻⁹, where the surrounding gas is shocked by the ejecta.

Our observations were part of a continuing effort^{3,10} to monitor by very-long-baseline interferometry (VLBI) the expansion of the radio supernova SN1993J. Here we discuss the data obtained on 26 September and 22 November 1993 with emphasis on the latter. We observed at 8.4 GHz with global VLBI arrays (Table 1). For the November 1993 data, after a few iterations of self-calibration using the California Institute of Technology VLBI software package¹¹, we obtained the image shown in Fig. 1 (using the CLEAN algorithm) with an excellent fit to the data (reduced $\chi^2 = 1$). An image obtained with the Maximum Entropy algorithm is essentially identical to the CLEAN image, showing clearly the same shell structure. For the September 1993 data, the source size and the available resolution of the array do not allow us to make a clear distinction between a shell and a disk.

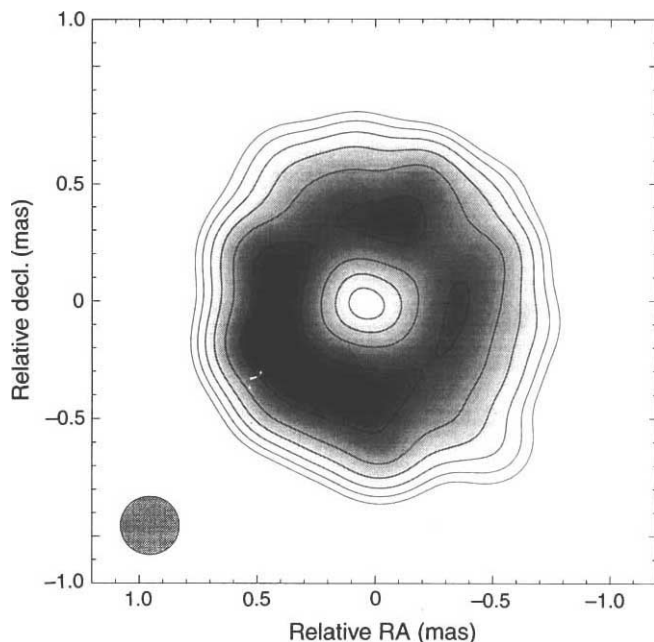


FIG. 1 Grey-scale and contour map of SN1993J from 8.4 GHz observations 239 days after explosion (RA, right ascension, decl., declination). As the visibility amplitudes are insensitive to a 180° rotation of the image, the (small) structure in the closure phases is required to resolve this ambiguity and favours slightly the enhanced emission being in the southeastern rather than the northwestern, part of the ring. The contours are at 2, 4, 8, 16, 32, 64 and 90% of the peak surface brightness. The gaussian beam used in the convolution of the δ -function-component model to obtain this CLEAN map is shown as a filled circle (at lower left) and has a full-width at half-maximum (FWHM) of 0.25 mas, half the value of the FWHM of the interferometer's main beam. 1 mas is equivalent to 5.45×10^{16} cm (0.054 light yr) for a distance of 3.63 Mpc.

The radio shell in Fig. 1 is almost spherically symmetric. Its detection and the relative uniformity of its brightness distribution exclude, at least for this supernova, models like the pulsar-driven model¹²⁻¹⁴. Other models^{6-9,15,16} predict supernova radio shells of approximate spherical symmetry.

But the brightness distribution around the shell is not totally uniform. The enhanced emission in the southeastern part of the ring is clearly implied by the data and is present in both the CLEAN and Maximum Entropy images. (This point is very relevant because these two algorithms have very different image deconvolution properties, with the former favouring point-like image structures and the latter extended image structures.) What could cause such a preferred direction? One possibility is the suggestion that the progenitor was a member of a binary system^{17,18}. Such a system could cause non-spherically symmetric density profiles of the pre-supernova circumstellar material, which would disrupt the spherical symmetry of the gas around the progenitor and affect in a similar manner the structure of the shock front responsible for the observed radio emission. Note that the progenitor of SN1987A was a single star and, despite that, asymmetrical emission from its otherwise spherically symmetric shell was observed¹⁹. But the densities of circumstellar matter, and therefore the emission levels and patterns, of SN1987A and SN1993J must be very different because of the different behaviour of these supernovae. In SN1987A, except for the initial radio flash, radio emission was not detected until more than two years after the explosion when the supernova was at a different, more advanced, stage of remnant formation than in SN1993J. In SN1993J, the radio emission was detectable two weeks after explosion and the light curves clearly show the optically thick and thin phases of a synchrotron emitter absorbed by a dense circumstellar medium²⁰.

The observed near-spherical symmetry of SN1993J is somewhat surprising given the asphericity implied by optical spectropolarimetric observations⁴, by line asymmetries⁵, and by suggestions that the progenitor was a member of a binary system^{17,18}. Perhaps the optical spectropolarimetric results⁴ from SN1993J (which were attributed to electron scattering in an aspherical geometry), as well as the line asymmetries, could be reinterpreted as being due to scattering from a spherically symmetric geometry with an asymmetric or non-uniform emission pattern. Or perhaps the observed polarization is due to phenomena other than electron scattering, and hence has nothing to do with the asphericity of the source.

In the standard model^{6-9,15,16} of a supernova the radio emission is assumed to be produced in a region of circumstellar gas shocked by the outgoing ejecta, thus giving rise to a 'mini-shell' emission, in contrast to the large-shell emission of historical supernova remnants like Cassiopeia. If the expansion rate is characterized by the parameter m in $\theta \propto t^m$ where θ indicates the source size and t the time elapsed since the explosion, then, using the standard model, we can determine the gas profile in the steep-gradient region of the shocked ejecta and thus perhaps derive some understanding of pre-explosion hydrodynamic conditions in the progenitor. Comparing our data in a consistent manner with earlier data we estimate $m = 0.90 \pm 0.05$, which implies a deceleration of the ejecta. The circumstellar material density profile has been found²⁰ to be proportional to $r^{-1.5}$, where r is the radial distance from the centre of explosion. Combining these results in the 'mini-shell' models^{6-8,15,16}, we infer that the density of the supernova ejecta has a very steep profile, proportional to $r^{-(17 \pm 5)}$.

Using our estimate for the radio-shell radius for 22 November 1993 (day 239 after explosion) and a zero-size at explosion, we obtain an average expansion rate for the shell radius of $2.43 \pm 0.15 \mu\text{s d}^{-1}$, rather lower than the estimate of $2.98 \pm 0.08 \mu\text{s d}^{-1}$ based on data from the first three months after explosion and using uniform-disk models³ for the source. Early radio observations did not have the necessary resolution to resolve the source (whether its structure was a shell or a disk) hence giving rise to this discrepancy. The value quoted for the standard error for our estimate of the average expansion rate is based on measurements of a set of our images, made with various sizes of the restoring beam, and is 40% larger than the statistical standard error from model fitting.

Given angular expansion rates and maximum gas expansion speeds, we can estimate^{3,21} the distance to M81. Previous estimates³ from analyses of optical spectra indicated that the maximum gas expansion speed was $18,000 \pm 1,000 \text{ km s}^{-1}$. But we note that a 'kink' in the H α absorption profile at $\sim 15,000 \text{ km s}^{-1}$, probably due to contamination from another line (perhaps Fe II 6,415 Å or possibly Si II 6,355 Å), make those estimates uncertain. Such a contamination line was observed earlier in SN1990E²². Because the removal of this possible contamination line from the observed profile cannot be done reliably at present, we believe that a more appropriate estimate is

TABLE 1 8.4 GHz observations and source parameters of SN1993J

Date of observation	Age* (d)	Measured total flux density† (mJy)	Circular disk diameter‡ (μas)	Shell outer diameter§ (μas)	Ratio of shell thickness to outer radius§
26 Sept. 1993	182	78 ± 4	478 ± 8		
22 Nov. 1993	239	59 ± 3		581 ± 35	0.3 ± 0.1

On 26 September 1993 the observations were made with the following antennas (diameters, affiliations and locations in parentheses): VLA (equivalent diameter 130 m, NRAO, near Socorro, New Mexico, USA); Effelsberg (100 m, MPIFR, Effelsberg, Germany); DSS63 (70 m, NASA, Robledo, Spain); DSS15 (34 m, NASA, Goldstone, California, USA); Medicina (32 m, CNR, Medicina, Italy). On 22 November 1993, instead of the DSS15 antenna, DSS14 (70 m, NASA, Goldstone, California, USA) was used. At each station a hydrogen-maser frequency standard was used to govern the local oscillator chain and the 'time-tagging' of the recorded data. The data were recorded with the Mark IIIA or compatible system, in mode B double speed, yielding bandwidths of 56 MHz. Right-hand circular polarization (IEEE convention) was recorded. The data were correlated at the Max-Planck-Institut für Radioastronomie, Bonn, Germany.

* Age of the supernova with respect to estimated explosion data of 28 March 1993.

† Obtained from observations with the VLA⁶.

‡ Estimate from model fitting; see text. Errors quoted are ~ 3 times the statistical standard deviations and include a 2% standard error in the amplitude-scale calibration for the shell.

§ Estimate from direct measurements on the image given in Fig. 1. See text.

$16,000 \pm 3,000 \text{ km s}^{-1}$. This estimate includes all known contributions to the uncertainty and is consistent with our inferences from other strong lines in the optical and ultraviolet that, unfortunately, have similar problems associated with their interpretation. Combining this result with our estimate of the average angular expansion rate and allowing for a 10% difference in the optical- and radio-shell growth rates²¹, we obtain $3.8 \pm 0.8 \text{ Mpc}$ for the distance to M81. This result is consistent with the independent value of $3.63 \pm 0.34 \text{ Mpc}$ (ref. 23) obtained from the Hubble Space Telescope observations of Cepheids in M81, which lends additional support to the modelling results.

At present we do not know how, if at all, the regions of enhanced emission within the shell are related to instability regions. The standard 'mini-shell' model does not make any predictions about the development of instabilities. To track the evolution of the remnant, continued VLBI observations at 8.4 GHz would be preferred. But such observations are now difficult to perform as the flux density at high frequencies has declined considerably, and the shell structure would be confused with the sidelobes of the VLBI point-spread function. Therefore observations at 5 GHz and 2.3 GHz are now underway. □

Received 28 June; accepted 11 November 1994.

- Canal, R., Isern, J. & Labay, J. A. *Rev. Astr. Astrophys.* **28**, 183–214 (1990).
- Ripero, J. & García, F. *IAU Circ. No.* 5731 (1993).
- Bartel, N. et al. *Nature* **368**, 610–613 (1994).
- Trammell, S. R., Hines, D. C. & Wheeler, J. C. *Astrophys. J.* **414**, L21–L24 (1993).
- Spyromilio, J. *Mon. Not. R. astr. Soc.* **266**, L61–L64 (1994).
- Chevalier, R. *Astrophys. J.* **259**, 302–310 (1982).
- Chevalier, R. *Astrophys. J.* **259**, L85–L89 (1982).
- Chevalier, R. *Astrophys. J.* **285**, L63–L66 (1984).
- Chevalier, R. in *Supernovae* (ed. Petschek, A. G.) 90–110 (Springer, New York, 1990).
- Marcaide, J. M. et al. *Astrophys. J.* **424**, L25–L27 (1994).
- Pearson, T. J. *Bull. Am. Astr. Soc.* **23**, 991–992 (1991).
- Shklovskii, I. S. *Pis'ma Astr. Zh.* **7**, 479–481 (1981); translated in *Soviet Astr. Lett.* **7**, 263–264 (1981).
- Pacini, F. & Salvati, M. *Astrophys. J.* **245**, L107–L108 (1981).
- Bandiera, R., Pacini, F. & Salvati, M. *Astrophys. J.* **285**, 134–140 (1984).
- Weiler, K. W., Sramek, R. A. & Panagia, N. *Astrophys. J.* **301**, 790–812 (1986).

- Weiler, K. W., Panagia, N. & Sramek, R. A. *Astrophys. J.* **364**, 611–625 (1990).
- Nomoto, K. et al. *Nature* **364**, 507–509 (1993).
- Podsiadlowski, Ph., Hsu, J. J. L., Joss, P. C. & Ross, R. R. *Nature* **364**, 509–511 (1993).
- Staveley-Smith, L. et al. *Nature* **366**, 136–138 (1993).
- Van Dyk, S. D., Weiler, K. W., Sramek, R. A., Rupen, M. P. & Panagia, N. *Astrophys. J. Lett.* **432**, L115–L118 (1994).
- Bartel, N. et al. *Nature* **318**, 25–30 (1985).
- Schmidt, B. P. et al. *Astr. J.* **105**, 2236–2250 (1993).
- Freedman, W. L. et al. *Astrophys. J.* **427**, 628–655 (1994).

ACKNOWLEDGEMENTS. We thank the many people that have made possible this research: B. Clark, M. Klein, R. Schwartz, P. Wolken and the staff at the participating observatories, at NASA/JPL, and at the Max-Planck-Institut für Radioastronomie, where all the data were correlated. We also thank A. P. Marscher, C. Masson, J. M. Moran and K. W. Weiler for comments on our manuscript. This research was partly supported by the Dirección General de Investigación Científica y Tecnológica (DGICYT) in Spain and by the US NSF in the US. NRAO is operated by Associated Universities, Inc., under a cooperative agreement with the US NSF; DSS14, DSS15, and DSS63 are operated by NASA/JPL.

Jet-like features near the nucleus of Chiron

J. L. Elliot^{*††}, C. B. Olkin^{*†}, E. W. Dunham[§],
C. H. Ford^{||}, D. K. Gilmore[¶], D. Kurtz[#],
D. Lazzaro^{**}, D. M. Rank[¶], P. Temi[¶],
R. M. Bandyopadhyay^{*}, J. Barroso^{**}, A. Barucci^{††},
A. S. Bosh[†], M. W. Bule[†], S. J. Bus^{*}, C. C. Dahn^{††},
D. W. Foryta^{§§}, W. B. Hubbard^{||}, D. F. Lopes^{**},
R. L. Marcialis^{||¶¶}, S. W. McDonald^{*}, R. L. Millis[†],
H. Reitsema^{##}, D. G. Schleicher[†], B. Sicardy^{††***},
R. P. S. Stone^{†††} & L. H. Wasserman[†]

CONSIDERED as a comet, the object 2060 Chiron is unusual in two respects: it exhibits outbursts at very large distances from the Sun¹⁻³, and its nucleus is much larger than that of any other known comet^{4,5}. It is, however, similar in size to the recently discovered Kuiper-belt objects⁶—a population of objects with orbits beyond Neptune, which are a possible source of short-period comets. This has led to the conjecture that Chiron is related to these objects, but its chaotic orbit has brought it much closer to the Sun⁷. Here we report observations of a recent stellar occultation by Chiron which permit the identification of several features associated with Chiron's coma. The observation of discrete, jet-like features provides evidence that the coma material originates from just a few, small active areas, rather than from uniform sublimation, and that the particles in at least one of these features have radii greater than 0.25 μm . The observations also suggest the presence of material in the plane of Chiron's orbit and are consistent with a gravitationally bound coma. Finally, the present data, and those from a previous occultation⁸, constrain the radius of Chiron to lie between 83 and 156 km.

The 9 March 1994 occultation of Ch08 (GSC248-01674)⁹ by Chiron was successfully observed with two telescopes: the 0.9-m telescope aboard NASA's Kuiper Airborne Observatory (KAO) flying near Recife, Brazil and the 0.5-m telescope at the South African Astronomical Observatory (SAAO) at Sutherland. We attempted observations with five other telescopes in Brazil. But their positions turned out to be far from the final predicted shadow path of Chiron, and all were clouded out. Instrumental parameters for the observations are specified in Table 1. Compared with its photometric history², Chiron was relatively faint at the time of our observations, with a standard magnitude¹⁰ $H_V = 6.6 \pm 0.1$.

We generated a light curve from the KAO optical data using the technique of point-spread-function (PSF) model fitting. The infrared light curve was produced by doing aperture photometry with a synthetic aperture three pixels in radius (image radii ranged between 1.0 and 2.0 pixels). The SAAO light curve was obtained directly from the recorded photomultiplier-tube integrations. Sections of the light curves that bracket Chiron's closest approach to the star are shown in Fig. 1. A sharp drop and recovery occurs in all light curves, which we shall denote as feature 1 (F1). This appears similar to a feature in the Palomar light curve for the earlier Chiron occultation—one

interpretation of which is a dust jet⁸. In addition, broader and shallower occultation features are particularly evident in the KAO optical light curve. To show the occultation features more clearly, shorter sections of the light curves are shown in Fig. 1, right panel.

In addition to F1 in the KAO optical light curve, we also see a broader feature (F2), that has a full-width at half-maximum (FWHM) of ~ 74 km, but less optical depth. Due to lower signal-to-noise ratio, F2 is not detectable in the infrared or SAAO light curves (although there is a hint of it in the former). Also, we note in the KAO optical light curve a low-level attenuation of the stellar signal that begins at ~ 517 s after the reference time and a similar attenuation between F1 and F2. Noting that F2 has a steeper slope on its leading edge than its trailing edge, one could interpret F2 as a symmetrical feature embedded in a low-level attenuation that ends at the trailing edge of F2. Without knowing how to apportion the optical depth into real, physical features, we have taken the liberty of labelling F3 as a feature that begins with the first low-level attenuation and ends at the trailing edge of F2. Finally, a feature labelled F4 is present in the KAO optical light curve. Its significance is 3σ (well below the detection limits of Earth-based imaging), and it occurs when the star aligns with Chiron's orbit plane, where material may be expected. Table 2 gives information about the optical depths and widths of the features.

To establish the geometrical relationship of these features in the light curves to the nucleus of Chiron, in Fig. 2 we have plotted a Chiron-centred sky frame¹¹ that shows the orbit plane, the track of the star from both stations, and the direction to the Sun. Because Chiron was only 15° from opposition, the largest component of the vector to the Sun is actually perpendicular to the plane of Fig. 2. The relative position of the star and Chiron's centre of light (assumed to be the centre of the nucleus) was determined from 40 astrometric frames taken with the 61-inch astrometric reflector at the US Naval Observatory, which yielded formal errors (1σ) of ± 16 km in right ascension and declination (the f, g coordinate systems in Fig. 2). The broad, low-optical-depth feature (F3) is nearly symmetrical around the closest approach time of the event, and F4 occurs near the orbit-plane crossing for the KAO (at 575 ± 5 s after 23:20:01 UT—the uncertainty is larger than the uncertainty in the closest approach owing to the oblique intersection of the apparent star path with the orbit plane).

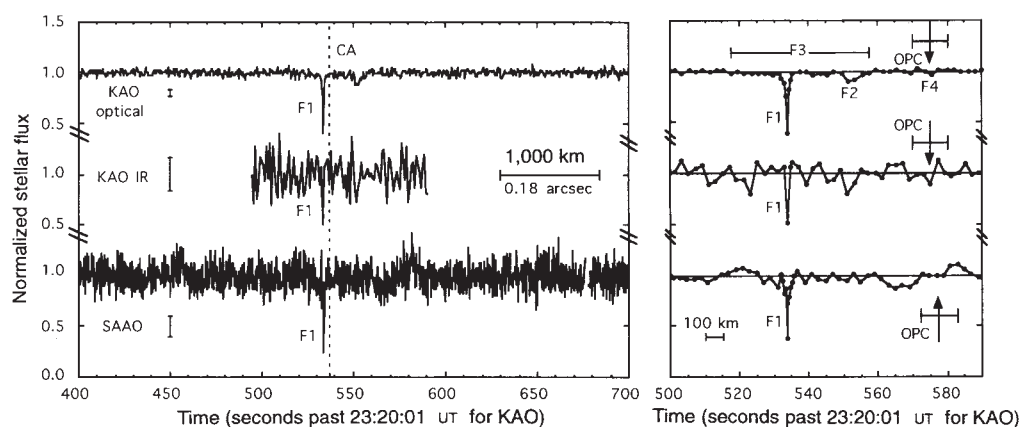
We can argue against the possibility that F1 is the signature of a small, sharp-edged body or that of a grazing occultation by the nucleus. Although a small body (or the nucleus) could produce a partial occultation, its characteristic dimensions would be of the order of the Fresnel diffraction scale $\sqrt{\lambda D}$, where λ is the wavelength of observation and D is the body-observer distance. These values are 0.9 and 1.6 km for the optical and infrared data respectively (the projected angular diameter of the star at Chiron's distance would be even less, only 0.1 km). The wings of F1 in the KAO optical light curve are much broader than this, ruling out a solid body or nuclear occultation (unless either of these were surrounded by a dust cloud that produced the wings).

Accepting that F1 is not a partial nuclear occultation, we can place a 3σ upper limit on Chiron's radius of 156 km from our closest approach distance of 108 ± 16 km. Compare this with the 83 km lower limit on Chiron's radius derived from the previous stellar occultation⁸ and radii ranging between 74 ± 11 km and 104 ± 10 km from thermal infrared observations⁵.

Other possibilities for F1 would be a suborbital arch, dust jet, or shell—the last two of which are seen in the comae of comets¹² and emanating from the nucleus of comet Halley^{13,14}. This interpretation would be consistent with the smooth appearance of the coma from ground-based imaging¹⁵⁻¹⁷, as all features we see would not be resolved in those data. The interpretation of F1 as caused by the outflow from a conical jet requires a spreading angle of $\sim 12^\circ$ for our nominal geometrical solution and a nuclear

* Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA. † Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA. ‡ Lowell Observatory, Flagstaff, Arizona 86001, USA. § NASA Ames Research Center, Space Science, MS 245-6, Moffett Field, California 94035-1000, USA. || SETI Institute, 2035 Landings Drive, Mountain View, California 94043, USA. ¶ University of California, Lick Observatory, Santa Cruz, California 95064, USA. ** University of Cape Town, Private Bag, Rondebosch 7700, Cape Town, South Africa. *** CNPq/Observatório Nacional, DAF, 20921 Rio de Janeiro, Brazil. †† Observatoire de Paris, 92195 Meudon Cédex Principal, France. ‡‡ US Naval Observatory, PO Box 1149, Flagstaff, Arizona 86002-1149, USA. §§ UFRP/Dep. Física, Centro Politécnico, 80000 Curitiba, Brazil. ||| Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. ¶¶ Pima Community College, 2202 West Anklam Road, Tucson, Arizona 85709, USA. ## Ball Aerospace, PO Box 1062, Boulder, Colorado 80306-1062, USA. *** Université Paris 6, UFR physique 924, 75252 Paris Cedex 05, France. ††† Lick Observatory, Mount Hamilton, California 95140, USA.

FIG. 1 Chiron occultation light curves. Portions of the stellar flux (normalized to 1.0 for the unocculted star) from Ch08 have been plotted versus ut for an interval that includes Chiron's closest approach to Ch08 (shown by the dotted line labelled "CA"). The left-hand panel is at full resolution. The right-hand panel includes only data near closest approach; to enhance the signal-to-noise ratio, data outside the sharp drop (feature 1, F1) have been averaged to 2.0 s. The segment of the light curves comprising F1 have been left at full resolution for the KAO data, whereas the SAAO data have been averaged to 0.4 s (two integration times and two deadtimes). In the left-hand panel the error bars are $\pm 1\sigma$, and the 1,000 km spatial scale at Chiron (for the KAO shadow velocity of 18.32 km s^{-1}) is approximately the resolution of the Hubble Space Telescope. The orbit-plane crossing (OPC) time is indicated in the right-hand panel. A sharp drop and recovery (F1) is evident in all three light curves. As the paths of the KAO and SAAO were separated in Chiron's shadow plane by only 1.4 km at closest approach (107.8 km for the KAO and 106.4 km for SAAO), we have assumed that F1 was



caused by the same material for both stations. We therefore corrected for the unknown timing offset in the SAAO data by retarding the time axis of the light curve by 309.9 s to align the sharp feature with that of the KAO. Near 680 s, there is a gap in the SAAO light curve, where a noise spike was removed. These spikes reached a flux level 100% above the Ch08 level and were removed before calculating the unocculted level for Ch08. See text for description of the features.

radius of 83 km (ref. 8), if we assume that the jet originates on the limb in the sky plane and from a small active area on the surface—consistent with the 1–2 km radii for active areas that would have been required to produce the visible coma at earlier times^{3,16–18}. Larger and smaller spreading angles are possible for other assumptions.

The ratio of extinction at infrared to that at optical wavelengths for F1 is formally 1.31 ± 0.65 (over equal integration intervals of 1.0 s). For the particle-size distributions found for comet Halley ($n^\alpha dn$, with α between -4 for particles with radii of $\sim 100 \mu\text{m}$ and -1.54 for those with radii $\sim 0.1 \mu\text{m}$)¹⁹, this ratio would be much less than one. Hence the extinction in F1 can not be caused by this type of particle-size distribution, but must be dominated by large particles. To obtain some quantitative estimate of the lower limit on the particle size, we use Mie theory

for spherical particles²⁰, and find that for various compositions from silicate to tholin to graphite these data are consistent with a lower limit of $0.25 \mu\text{m}$ (ref. 21) on the radius of particles. So most of the extinction we measure must be contributed by particles larger than this, but somewhat different limits would apply for non-spherical particles, with indices of refraction different from our assumption²⁰. This conclusion for F1 shows a particle population differing from that measured²² by the particle detectors on the Halley spacecraft, which found a significant population of particles with radii smaller than $0.25 \mu\text{m}$. The material probed here, however, is from a different object, much further from the Sun, and sampled much closer to the nucleus. An upper limit to the radii of particles that can be lifted from the nucleus by escaping gas has been estimated¹⁷ at $100 \mu\text{m}$ for CO and $10 \mu\text{m}$ for CO_2 .

TABLE 1 Observations of the Ch08 occultation by Chiron

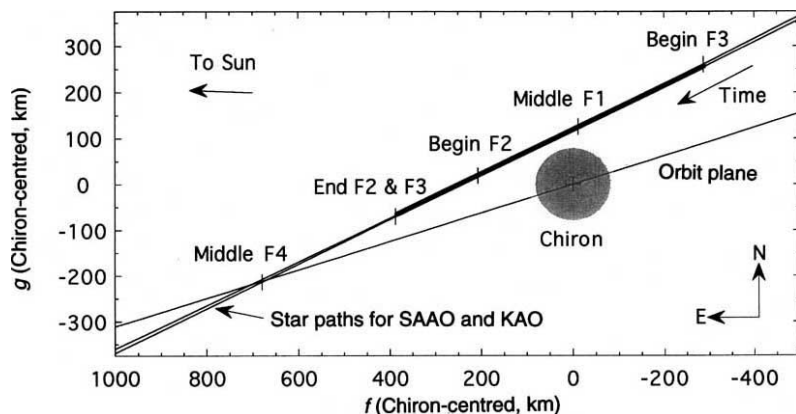
Telescope (diameter)	Detector	Filter	Recording interval (ut)	Number of integrations	Integration time (s)	S/N (for 1 s)	S/N (for 10 km)
KAO (0.9 m)	CCD	None*	23:20:01–23:36:40.5	2,000	0.5	52	38
KAO (0.9 m)	HgCdTe array	K	23:20:01–23:36:40.5	1,000	1.0	5.8	4.3
SAAO (0.5 m)	PMT	V	22:58:53–23:42:40†	13,266	0.1†	20	15

The occulted star is the brighter of two stars that have the combined photometric measurements: $V=11.95 \text{ mag}$, $(B-V)=0.92$, $(V-R)=0.42$, $(V-I)=0.86$, $(V-K)=1.9$. The fainter companion star lies at a distance of 3.27 arcsec from the brighter one, at a position angle of 19.6° , and it has magnitude differences with Ch08 of 2.15 in the K band and 3.83 in a broadband red filter (6,600–8,300 Å), used by the US Naval Observatory. The KAO optical data were recorded with a high-speed CCD (charge-coupled device) photometer²⁴, using two 60×60 pixel subframes (0.98 arcsec per pixel): one subframe contained the merged images of Chiron, Ch08 and the faint companion star; the other contained a field star of brightness comparable to Ch08 and 4.38 arcmin away. A continuous series of 2,000 subframe pairs were recorded, and the timing of the series was related to ut with a GPS clock. A beamsplitter enabled simultaneous infrared observations to be carried out with a NICMOS HgCdTe array (128×128 pixels, ~ 1.4 arcsec per pixel) that was read out once per second, synchronously with the optical data recording. The SAAO data were recorded with a photomultiplier tube (PMT) at an integration time of 0.1 s, through an aperture 30 arcsec in diameter. Reference to the standard star HR4064 showed that the night was photometric at SAAO. The deadline between integrations was usually 0.1 s, but, sometimes it was 0.0 s, which made it impossible to accurately relate the recording times of these data to ut . The signal-to-noise ratio S/N in the next to last column is defined as the ratio of the unocculted flux of the star to the r.m.s. noise, each integrated for one second. The last column contains the analogous ratio for a 10 km interval in the sky-plane.

* The unfiltered CCD response and stellar spectrum combine for a mean wavelength of $\sim 0.68 \mu\text{m}$.

† The end time of the SAAO data is not precisely known. The value here was calculated for a deadline of 0.1 s between integrations for 98% of the data, and no deadline between the remaining 2% of the integrations (as the deadline between successive integrations was 0.1 s about 98% of the time and 0.0 s for the remaining 2%).

FIG. 2 The Chiron-centred sky-plane. Chiron's shadow at Earth is indicated in the (f, g) coordinate system¹¹, in which f is parallel to right ascension and g is parallel to declination. The orbit plane of Chiron has been plotted for reference. Along the paths probed by ChO8 as seen from the KAO and SAAO, relevant times for light-curve features are indicated, and time increases to the south-east. Note that the occultation probed the orbit plane ahead of Chiron. Along the apparent path of the star, we have indicated the positions of the features derived from the KAO optical light curve. Coincidentally both stations probed nearly the same region near Chiron. The uncertainty in the closest approach time is the geometrical error divided by the shadow velocity (~ 0.9 s). Chiron was 15° from opposition at the occultation time, so the vector to the Sun is mostly out of the plane of the figure. A Chiron radius of 83 km has been assumed⁸ (indicated by the shaded circle). For the KAO, the occultation shadow probes the orbit plane at 575 s after 23:20:01 UT, and there is a 3σ occultation feature evident in the light curve at this time (F4). If F1 and F2 are caused by jets originating from the limb of the nucleus nearest to where these features were detected, then extending lines from F1 and F2 to Chiron's centre, we see that F2 would be on the side of the nucleus facing the Sun (consistent with the jet



hypothesis) and F1 would be on the sunrise or sunset limb (marginally consistent with the jet hypothesis). Owing to projection effects, however, either feature could be connected to a sunlit or dark portion of the nucleus. Note that F3 is nearly symmetrical with respect to Chiron, as would be expected for a bound coma. At closest approach, Chiron was 8.83 AU from the Sun and 7.87 AU from the Earth.

If one interprets F1 as a jet, it is natural to interpret F2 as a jet also. Being probed further from the nucleus than F1, it has spread into a broader morphology. In fact, it is possible that F1 and F2 are caused by a single structure—shaped as a spiral or suborbital arch—that has been probed twice. In order to test this hypothesis, one would need to know the rotation of the nucleus. Although the period has been measured quite accurately^{2,3,17}, the direction of the rotation pole²³, estimated from the imaged coma morphology, disagrees somewhat with constraints from photometry². Because we are seeing these features projected onto a plane, however, their distances from the nucleus in this plane are only lower limits on their true distances. Furthermore, the vector to the Sun points mostly out of the sky plane, so that either F1 or F2 could be emanating from a sunlit or dark side of the nucleus. Despite these uncertainties, we presently feel that F1 and F2 are probably caused by two distinct dust jets from the nucleus.

Although F3 is much broader, it could possibly be a jet, shell, or other shape¹²—seen much further from the nucleus than F1 or F2, but appearing nearly symmetrical with respect to the nucleus owing to projection effects. Another, and perhaps more plausible, explanation for F3 is that it is a symmetrical cloud of dust around the nucleus. If this material is flowing out uniformly at the speed of sound for CO (160 m s^{-1})¹⁷, then the material rate of loss (for particles $10 \mu\text{m}$ in radius and density 1 g cm^{-3}) would be several orders of magnitude larger than that needed to account for the extended coma (in 1990, when Chiron was at least 0.5 mag brighter in H_V)^{16,17,23}. Therefore we believe that the material causing F3 is not flowing outwards from the nucleus at the speed of sound, but is possibly gravitationally bound to the nucleus—a bound coma^{16,17}. Such a feature would be unique to Chiron because it is the only known comet large enough for gravity to hold dust in bound orbits for periods ranging from tens of days¹⁸ to a year¹⁶. If F3 is a manifestation of a bound

TABLE 2 Material detected near Chiron

Feature	Light curve	Midtime (s after 23:20:01 UT)	Distance from Chiron's centre (km)*	Averaging interval (km)	Maximum optical depth measured	FWHM (km)	Integrated optical depth (km)†	Morphology
1	KAO, optical KAO, infrared SAAO	533.7	126 ± 16	9.2	$0.92 \pm 0.07^\ddagger$	5.5–9.2	13.2 ± 1.3	Narrow, deep, symmetrical
				18.3	0.68 ± 0.34	9.0–18.3	12.4 ± 6.2	
				7.3	0.98 ± 0.46	4.5–7.3	7.2 ± 3.4	
2	KAO, optical	552.4	295 ± 16	Model fit	0.11 ± 0.02	74 $\pm$ 15	4.1 ± 1.3	Asymmetric?
3	KAO, optical	537.4	108 ± 16	Model fit	0.027 ± 0.006	513 $\pm$ 169	21.6 ± 5.0	Broad, flat
4	KAO, optical	575.7	712 ± 16	Model fit	0.07 ± 0.04	12 $\pm$ 10	1.6 ± 0.5	3σ

The averaging interval is equal to the product of the apparent shadow velocity from the site and the integration interval for the data, and the maximum measured optical depth is derived from the integration having the lowest signal level. The differences between the integrated optical depth in the SAAO and KAO optical light curves can be attributed to one or more of the following: (1) the core of feature (F1) is not resolved in the KAO optical light curve, because the integration time is longer than for SAAO, (2) the duty cycle of the SAAO data recording is 50%, whereas that for the KAO data is 100%, or (3) F1 changed during the 310 s that elapsed between the recording of F1 at the two stations. The maximum optical depths and widths of F2, F3 and F4 were determined from fitting a lorentzian model.

* Distance of feature from Chiron's centre probed at midtime (column 3).

† The integrated optical depths depend on how one chooses to separate the features. For features 1 and 2 (F1 and F2) of the KAO optical light curve, the integrated optical depth was determined by taking the logarithm of the summed normalized stellar flux for all points in the feature (after correcting for the optical depth per point from F3 of 0.027). This optical depth was then converted to kilometres by multiplying by the velocity and integration time. For F3 the integrated optical depth was determined by model fitting (to include the optical depth contribution when this feature is embedded in the other features). The integrated optical depth for F4 was also determined by model fitting. Finally, the integrated optical depths for the KAO infrared and the SAAO data were calculated by multiplying the maximum optical depth by the averaging interval, with the SAAO data averaged over two points (for a time interval of 0.4 s, including the deadtime).

‡ Optical depth averaged over 18.3 km of projected width for feature 1 is 0.52 ± 0.03 . This average should be used when comparing with the average optical depth in the infrared. The extinction ratio (infrared to optical) is 1.31 ± 0.65 .

coma, it may be only a small part, with the remainder having too low an optical depth to be detectable in these data. □

Received 15 August; accepted 24 November 1994.

- Tholen, D. J., Hartmann, W. K. & Cruikshank, D. P. *IAU Circ. No.* 4554 (1988).
- Marcialis, R. L. & Buratti, B. J. *Icarus* **104**, 234–243 (1993).
- Bus, S. J., Howell, E., Harris, A. W. & Hewitt, A. V. *Icarus* **77**, 223–238 (1989).
- Sykes, M. V. & Walker, R. G. *Science* **251**, 777–780 (1991).
- Campins, H. *et al. Astr. J.* **108**, 2318–2322 (1994).
- Jewitt, D. & Luu, J. *Nature* **362**, 730–732 (1993).
- Scholl, H. *Icarus* **40**, 345–349 (1979).
- Buie, M. W. *IAU Circ. No.* 5898 (1993).
- Bus, S. J., Wasserman, L. H. & Elliot, J. L. *Astr. J.* **107**, 1814–1824 (1994).
- Howell, E. *et al. in Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) (Univ. Arizona Press, Tucson, 1989).
- Elliot, J. L. *et al. Astr. J.* **106**, 2544–2572 (1993).
- Sekanina, Z. *in Comets in the Post-Halley Era* (eds Newburn, R. L., Neugebauer, M. & Rahe, J.) 769–823 (Kluwer, Dordrecht, 1991).
- Sagdeev, R. Z. *et al. Nature* **321**, 262–266 (1986).
- Sekanina, Z. & Larson, S. M. *Nature* **321**, 357–361 (1986).
- West, R. M. *Astr. Astrophys.* **241**, 635–645 (1991).
- Meech, K. J. & Belton, M. J. S. *Astr. J.* **100**, 1323–1338 (1990).
- Luu, J. X. & Jewitt, D. C. *Astr. J.* **100**, 913–932 (1990).
- Stern, S. A., Jackson, A. A. & Boice, D. C. *Astr. J.* **107**, 765–771 (1994).
- Lamy, P. L., Grün, E. & Perrin, J. M. *Astr. Astrophys.* **187**, 767–773 (1987).
- van de Hulst, H. C. *Light Scattering by Small Particles* (Dover, New York, 1981).
- Olkin, C. B. *et al. in Proc. Airborne Astronomy Symp. on the Galactic Ecosystem: From Gas to Stars to Dust* (eds Haas, M. R., Davidson, J. A. & Erickson, E. F.) (Astr. Soc. Pacif., San Francisco, 1994).
- McDonnell, J. A. M., Lamy, P. L. & Pankiewicz, G. S. *in Comets in the Post-Halley Era* (eds Newburn, R. L., Neugebauer, M. & Rahe, J.) 1043–1073 (Kluwer, Dordrecht, 1991).
- Fulle, M. *Astr. Astrophys.* **282**, 980–988 (1994).
- Buie, M. W. *et al. Bull. Am. Astro. Soc.* **25**, 1115 (1993).

ACKNOWLEDGEMENTS. We thank the entire KAO crew for their outstanding work, especially mission director J. McClenahan and navigator R. Morrison; we also thank the members of the Brazilian and US governments who worked out agreements that allowed KAO to use Brazilian air space and airport accommodation on short notice. The NICMOSII FPA used in these observations was donated by Rockwell International Science Centre. This work was supported in part by NASA.

Graft copolymers that exhibit temperature-induced phase transitions over a wide range of pH

Guohua Chen & Allan S. Hoffman*

Center for Bioengineering, FL-20, University of Washington, Seattle, Washington 98195, USA

THERE are many potential applications of 'intelligent' aqueous polymer systems^{1–8} in medicine, biotechnology, industry and in environmental problems^{9–13}. Many of these polymer systems undergo reversible phase transitions—for example, abrupt changes in volume—in response to external stimuli such as temperature, pH or the nature of the solvent. Most of the polymers studied previously are responsive to only one kind of stimulus. But for some applications, independent responsiveness to several factors, such as temperature and pH, may be required. Here we describe a polymer that undergoes marked solubility changes in water in response to temperature and/or pH changes. The polymer is prepared by grafting temperature-sensitive side chains onto a pH-sensitive backbone. We also find that block copolymers, in which the temperature- and pH-sensitive units alternate along the chain, show similar behaviour.

When the temperature of a polymer solution is raised above the lower critical solution temperature (LCST) or cloud point (CP), the polymer will phase-separate¹⁴: hydrophobic groups in the polymer form insoluble aggregates, turning the solution cloudy. We have used this principle previously¹⁵ to prepare a copolymer hydrogel which exhibits temperature-sensitive

swelling-deswelling changes over a limited pH range. This gel is composed of a random copolymer of monomers that are pH-sensitive (acrylic acid, AAc) and temperature-sensitive (*N*-isopropyl acrylamide, NIPAAm). For a composition of less than 10 mol% AAc, the gel exhibits a CP at pH 7.4. For higher AAc content the CP disappears, because then the AAc components (which are ionized at pH 7.4) convey sufficient solubility to offset the aggregation of the hydrophobic temperature-sensitive components. High molecular weight PAAc is a well-known bioadhesive polymer, which 'sticks' to the hydrated mucosal cells coating the eye, nose, mouth, lungs, gastrointestinal tract, vagina and anus. In order to prolong the residence time of a drug delivery vehicle in contact with such mucosal surfaces, PAAc is often incorporated into a delivery formulation. Further, if it is also desirable to slow the rate of drug release from the bioadhesive formulation, a more hydrophobic component, such as the temperature-sensitive polymer, may be added. Random copolymers of AAc and NIPAAm will not work, because of the loss of the temperature sensitivity when the content (that is, molecular weight) of the AAc component is increased to the point where bioadhesive properties are obtained. Physical mixtures of these two types of polymers are also possible, but they tend to separate physically and release drug too rapidly. Thus, the graft copolymer structure, which combines both behaviours in a single molecule and does not permit physical separation, is most effective. This is the rationale for the work described here.

Copolymers¹⁶ and interpenetrating networks¹⁷ of two polymers that form hydrogen-bonded complexes involving carboxylic acid groups may also exhibit temperature sensitivity, but this usually disappears when the pH is raised above the pK of the acid groups, ionizing them to carboxylate. For example, Klier *et al.*¹⁶ prepared hydrogels of graft copolymers of poly(ethyleneglycol) (PEG) on a poly(methacrylic acid) (PMAAc) backbone, which form hydrogen-bonded complexes between –O– and –COOH groups at low pH. Although these gels gradually shrank as temperature was raised at pH 4, this temperature-sensitivity would not occur at pH 7.4 where the gels were highly swollen owing to ionization of the carboxyl groups and the consequent disruption of the complex¹⁶.

To obtain polymers that retain their temperature-induced transition over broad and useful compositional and pH ranges, especially at physiological conditions of pH 7.4 in a saline solution, we have synthesized graft copolymers composed of side-chains of a temperature-sensitive polymer (PNIPAAm) grafted onto a pH-sensitive backbone polymer (PAAc). We have used two methods for synthesizing the NIPAAm-g-AAc graft copolymers (Fig. 1*a–d*). In Fig. 1*e* we illustrate possible hydrogen-bonding interactions between the grafted NIPAAm and backbone AAc groups; these hydrogen-bonded structures are similar to those proposed to exist between carboxyl and amide groups in interpenetrating networks of PAAc and an acrylamide copolymer¹⁷.

In Fig. 2 we plot the effect of AAc content on CP for random and graft copolymers of NIPAAm and AAc at pH 7.4 and 4.0; that is, above and below the pK of AAc, respectively. It can be seen that the CP of the random copolymer is always higher than that of the PNIPAAm homopolymer and rapidly rises as the AAc content increases, especially at pH 7.4. In contrast, the graft copolymers show a constant CP at either pH, independent of AAc content over a wide range of compositions. Their LCST may rise only when the composition of the graft approaches 100% AAc (assuming that the few PNIPAAm graft chains present will still phase-separate individually, but without any visible evidence of turbidity in the solution). This constancy of CP over a broad composition range (especially at pH 7.4) is a clear demonstration of the significant differences in behaviour between the random and graft copolymers studied here, as well as between our graft copolymers and those of Klier *et al.*¹⁶ or the interpenetrating networks of Katono *et al.*¹⁷.

* To whom correspondence should be addressed.

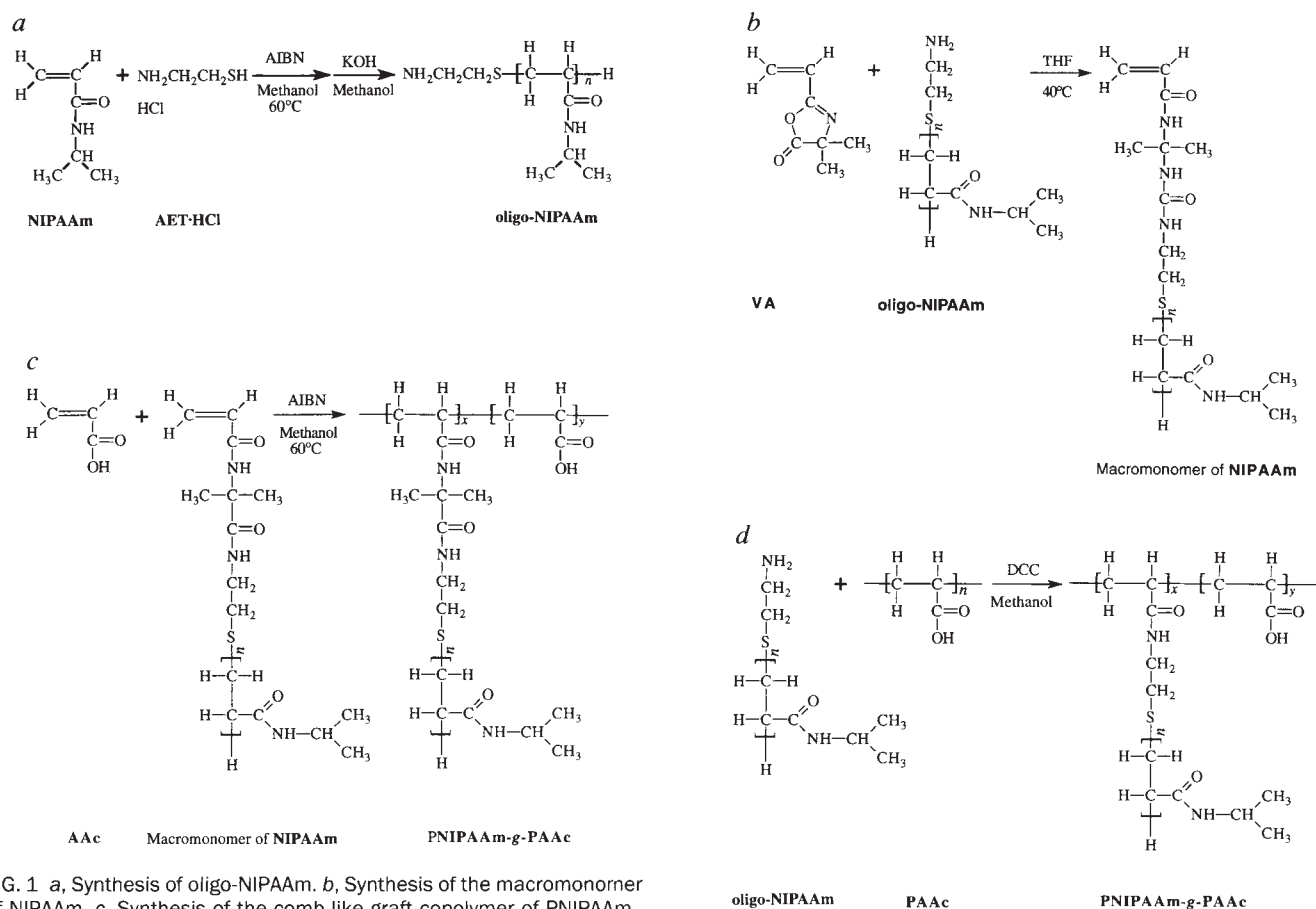


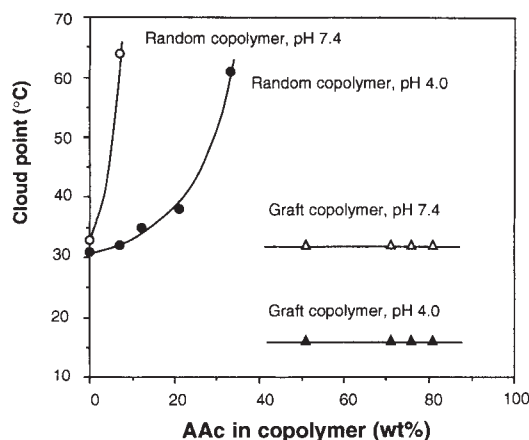
FIG. 1 *a*, Synthesis of oligo-NIPAAm. *b*, Synthesis of the macromonomer of NIPAAm. *c*, Synthesis of the comb-like graft copolymer of PNIPAAm-g-PAAc using the macromonomer copolymerization method. *d*, As *c* using the coupling method. *e*, Proposed H-bonding between NIPAAm graft chain and PAAc backbone.

METHODS. *a*, Oligo-NIPAAm with a terminal amino group was synthesized by free-radical polymerization of NIPAAm (11.3 g, 100 mmol) in methanol (40 ml) at 60 °C for 20 h using 2,2'-azobisisobutyronitrile (AIBN; 0.164 g, 1 mmol) and 2-aminoethanethiol hydrochloride (AET·HCl; 0.904 g, 8 mmol) as initiator and chain transfer reagent, respectively. The oligomer was obtained by precipitating the reaction solution into diethyl ether. The number average molecular weight of the oligomer was determined by titration to be 2,200. *b*, The macromonomer of NIPAAm was prepared by reacting 5.0 g (2.27 mmol) of amino-terminated oligo-NIPAAm with vinyl azlactone (VA; 0.944 g, 6.79 mmol) in 120 ml of dry tetrahydrofuran (THF) at 40 °C for 16 h. To synthesize the comb-like graft copolymers, two methods were used. *c*, The macromonomer of NIPAAm was copolymerized with AAc in methanol (10% w/v) using AIBN as initiator, at 60 °C for 1.5 h, with weight ratios of macro-

monomer/AAc of 20/80 and 40/60; *d*, The amino terminal group of the oligoNIPAAm was reacted with the carboxyl groups of PAAc in methanol (2% w/v) using dicyclohexyl carbodiimide (DCC) as the activation reagent at room temperature for 24 h, using weight ratios of oligo-NIPAAm/PAAc of 20/80, 25/75, 30/70 and 50/50.

FIG. 2 The phase-separation temperatures, or cloud points (CPs), of random copolymers of NIPAAm and AAc, and of comb-like graft copolymers of PNIPAAm-g-PAAc plotted against the copolymer AAc content at pH 4.0 and 7.4.

METHODS. The CPs of the copolymers were measured by spectrophotometric determination of the turbidity (absorbance at 500 nm) of the copolymer solution in buffer (2.0 mg ml⁻¹ in citric-phosphate buffer). The CP was determined at 10% absorbance. The random copolymers of NIPAAm and AAc were prepared by radical copolymerization of NIPAAm and AAc in methanol at 60 °C using AIBN as initiator.



The behaviours of the random and graft copolymers at the two pHs studied are also dramatically different. The higher CPs of the random copolymers at higher AAc contents at pH 4.0 are simply due to the hydrophilic character of the AAc monomer units along the backbone. This effect is known to raise the CP of temperature-sensitive random copolymers^{18,19}. At this same low pH, the graft copolymers show much lower CPs than homo-PNIPAAm. This is probably due to the efficient formation of hydrogen bonds between relatively long sequences of PNIPAAm and PAAc—as suggested in Fig. 1e and as shown for acrylamide and PAAc interpenetrating networks¹⁷—which could not form in the random copolymer. The hydrogen bonding interferes with the access of water molecules to the NIPAAm amide groups, thus rendering the graft chains of PNIPAAm more hydrophobic, lowering their phase-separation temperature. Based on this mechanism, the CP should be insensitive to the AAc content of the graft copolymer, as observed (Fig. 2). At pH 7.4, the $-\text{COOH}$ groups are ionized and the hydrogen bonds are disrupted. At this pH, the pendant PNIPAAm graft chains have the same CP as homo-PNIPAAm.

Figure 3 shows light absorbance versus temperature for 0.2% solutions of the 50% PNIPAAm-g-PAAc graft copolymer at pHs between those shown in Fig. 2. Distinct CPs are observed at all pHs studied, and the transitions tend to be sharper at the lower pHs. These CPs are shown as a function of pH in Fig. 4a: four different regions (A–D) are identified. It can be seen that the phase-separation temperature drops sharply as the pH is decreased below the pK of the backbone PAAc. The existence of a CP maximum is not significant. The slight decrease in CP at higher pH is probably due to the slight increase in ionic strength of the solution at the higher pHs²⁰.

Hypothetical conformations of the graft copolymers in the regions A–D are suggested in Fig. 4b. In region A, both the PAAc and PNIPAAm components are at their most soluble (high pH and low temperature) and the sketch of the expanded conformation of the graft copolymer illustrates this. As temperature is raised through the CP at constant pH (to region B), the PNIPAAm chains independently phase-separate, presumably mostly owing to the release of bound water as is the case for homo-PNIPAAm chains in solution¹⁴. The hydrophobic character of the precipitating chains may cause them to bond together, leading to the clouding of the solution, despite the highly hydrophilic and ionized backbone PAAc. The actual physical state of the 'precipitate' is not well-defined; the sketch of the conformation found in region B (Fig. 4b) is only schematic. A random copolymer of the same composition remains soluble over the entire temperature range, as can be deduced from Fig. 2.

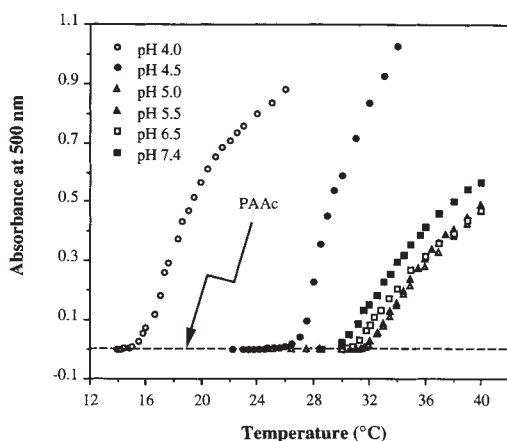


FIG. 3 The absorbance of the 50 wt% NIPAAm graft copolymer in 0.2 wt% solutions (in citric-phosphate buffer) versus temperature at various pH values.

As pH of the solution is lowered at constant temperature, the PAAc backbone chain will become less ionized and its expanded molecular coil should begin to shrink, but still remain in solution, as in region D (Fig. 4). In this state, some intramolecular hydrogen bonding between the grafted PNIPAAm chains and non-ionized sequences along the PAAc backbone may occur, as described above and shown in Fig. 1e. As pH continues to decrease from region A to D, the formation of hydrogen-bonded sequences will become increasingly probable. Eventually, a critical level of hydrophobicity of the grafted PNIPAAm chains will be reached at which they will precipitate, even at temperatures well below the CP of homo-PNIPAAm.

These graft copolymers represent a new family of 'hybrid intelligent' polymers. Furthermore, they may be fabricated in the form of soluble polymers, gel-coated surfaces or free-standing

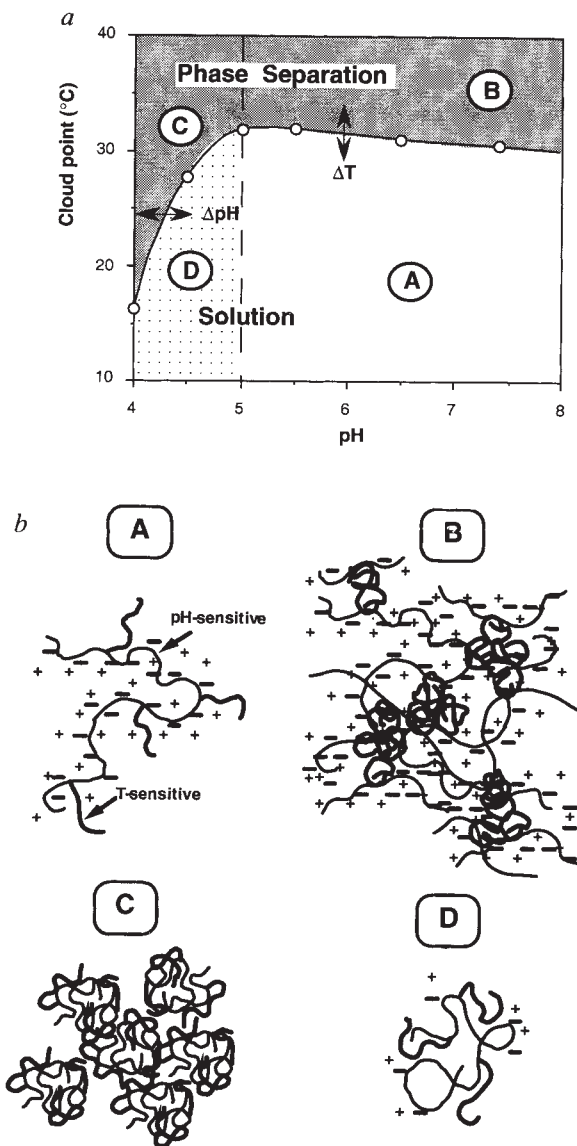


FIG. 4 a, The CPs of the graft copolymer (50 wt% NIPAAm) versus pH, showing four regions A–D. Two of the regions (B and C) represent conditions where the graft copolymer has phase-separated, whereas A and D are regions where the copolymer remains in solution. Arrows indicate reversible phase separations induced between regions A and B by a small change in temperature (ΔT) at constant pH, or between regions C and D by a small change in pH (ΔpH) at constant temperature. The dashed vertical line is placed at a pH near the pK of the PAAc, separating regions of relatively high and relatively low extents of ionization. b, Sketches of the molecular conformations in the four regions shown in a.

hydrogels. Block copolymer structures, including coating and hydrogel forms, are also possible. In addition, graft or block copolymers having components with sensitivities to different stimuli could become even more 'intelligent' by combining them with a variety of biomolecules. Applications of these copolymer-biomolecule systems include drug delivery, diagnostics, separations, cell culture and bioreactions. In addition to physical mixtures, if biomolecules are covalently conjugated onto some of the block or graft polymer chains, a 'hybrid intelligent' copolymer with amplified biological activity would be possible. □

Received 10 August; accepted 14 November 1994.

1. Hoffman, A. S. *Mater. Res. Soc. Bull.* **XVI**, 42–46 (1991).
2. Dusek, K. (ed.) *Responsive Gels: Volume Transitions II* (Springer, Berlin, 1993).
3. Tanaka, T. *Polymer* **20**, 1404–1412 (1979).

4. Kishi, R., Hara, M., Sawahata, K. & Osada, Y. in *Polymer Gels* (eds DeRossi, D. et al.) 205–220 (Plenum, New York, 1991).
5. Hoffman, A. S. *J. Control. Rel.* **6**, 297–305 (1987).
6. Freitas, R. F. S. & Cussler, E. L. *Chem. Engng Sci.* **42**, 97–103 (1987).
7. Okano, T., Bae, Y. H., Jacobs, H. & Kim, S. W. *J. Control. Rel.* **11**, 255–265 (1990).
8. Kungwachakun, D. & Irie, M. *Makromol. Chem., Rapid Commun.* **9**, 243–246 (1988).
9. Monji, N. & Hoffman, A. S. *Appl. Biochem. Biotech.* **14**, 107–120 (1987).
10. Chen, G. H. & Hoffman, A. S. *Bioconjugate Chem.* **4**, 509–514 (1993).
11. Bae, Y. H., Okano, T. & Kim, S. W. *Makromol. Chem., Rapid Commun.* **9**, 185–189 (1988).
12. Dong, L. C. & Hoffman, A. S. *J. Control. Rel.* **13**, 21–31 (1990).
13. Suzuki, M. in *Polymer Gels* (eds DeRossi, D. et al.) 221–236 (Plenum, New York, 1991).
14. Heskins, M. & Guillet, J. E. *J. Macromol. Sci., Chem.* **A2**, 1441–1455 (1968).
15. Dong, L. C. & Hoffman, A. S. *J. Control. Rel.* **15**, 141–152 (1991).
16. Klier, J., Scranton, A. B. & Peppas, N. A. *Macromolecules* **23**, 4944–4949 (1990).
17. Katono, J. et al. *J. Control. Rel.* **16**, 215–228 (1991).
18. Taylor, L. D. & Cerankowski, L. D. *J. Polym. Sci.* **A8**, 3405–3415 (1970).
19. Priest, J. H., Murray, S. L., Nelson, J. R. & Hoffman, A. S. *Am. Chem. Soc. Symp. Ser.* **350**, 255–264 (1987).
20. Park, T. G. & Hoffman, A. S. *Macromolecules* **26**, 5045–5048 (1993).

ACKNOWLEDGEMENTS. This work was supported by the Alcon Corporation.

Dominant influence of atmospheric circulation on snow accumulation in Greenland over the past 18,000 years

W. R. Kapsner*, R. B. Alley*, C. A. Shuman*, S. Anandkrishnan* & P. M. Grootes†

* Earth System Science Center and Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

† Quaternary Isotope Laboratory, University of Washington, Seattle, Washington 98195, USA

PROJECTIONS of sea-level rise due to greenhouse warming often involve the assumption that increased water vapour pressure will enhance snow accumulation in cold regions of ice sheets, partially offsetting the increased melting of low-latitude and low-altitude ice^{1–3}. To test whether this has been true in the past, we compare accumulation rates⁴ and temperatures derived from the oxygen isotope composition⁵ of ice in the deep core obtained by the Greenland Ice Sheet Project II (GISP2). We find that atmospheric circulation, not temperature, seems to have been the primary control on snow accumulation in central Greenland over the past 18,000 years. During both warm (Holocene) and cold (Younger Dryas, Last Glacial Maximum) climate regimes, the sensitivity of accumulation to temperature changes is less than expected if accumulation is controlled thermodynamically by the ability of warmer air to deliver more moisture. During transitions between warm and cold climate states, in contrast, accumulation varies more than can be explained in purely thermodynamic terms, probably because of changes in storm tracks. Thus, in a world warmed by the greenhouse effect, circulation changes may be more important than direct temperature effects in determining snow accumulation in Greenland and its contribution to sea-level change.

The GISP2 core was dated by counting annual layers^{4,6,7,35}, with an estimated accuracy of 1–2% over century-length time-scales in the Holocene period, and probably better than 5% accuracy in the late glacial⁴. After making ice-flow corrections⁸, we can determine the snow accumulation rates⁴. We restrict our analysis to the upper 60% of the ice sheet where ice flow is simple⁹ and ice-flow corrections are unlikely to add significant error to estimates of accumulation-rate changes over decades or centuries.

Temperatures were estimated from the oxygen isotope composition of ice⁵ $\delta^{18}\text{O}$ (‰), using a calibration derived from com-

parison to a borehole temperature (T) profile: $T(\text{K}) = [(\delta^{18}\text{O} + 18.2)/0.53] + 273$ (refs 10, 11), which is not significantly different from the spatial-gradient calibration for central Greenland¹². Broad agreement between isotopic and borehole palaeothermometers over the last glacial cycle¹³, and close agreement over the most recent millennium^{10,11}, show that this is a useful palaeothermometer. We show data from the latter part of the Holocene and from the deglacial transition. Isotope data from certain portions of the Holocene have a small offset due to fractionation during storage before analysis; although this offset can be corrected for, inclusion of the corrected data does not affect our main conclusions and we therefore do not include these data.

Three lines of evidence suggest that the slope of this calibration (0.53‰ per °C) might be significantly larger only at the major climate transitions. First, changes in deuterium excess across the Younger Dryas termination indicate the temperature change of the vapour source area^{14,15}. Second, general circulation model simulations of the isotopic composition of precipitation suggest that circulation changes affect the relative importance of different moisture sources and thus the isotopic composition of Greenland snow¹⁶. Last, simulations of the Younger Dryas cooling using general circulation models yield a smaller temperature change in central Greenland¹⁷ than that obtained from the isotopic shift and the usual slope of the calibration¹². In addition, any possible effect of ice-age dustiness on isotopic composition through changes in cloud supercooling would also increase the slope of the calibration at the major climate transitions.

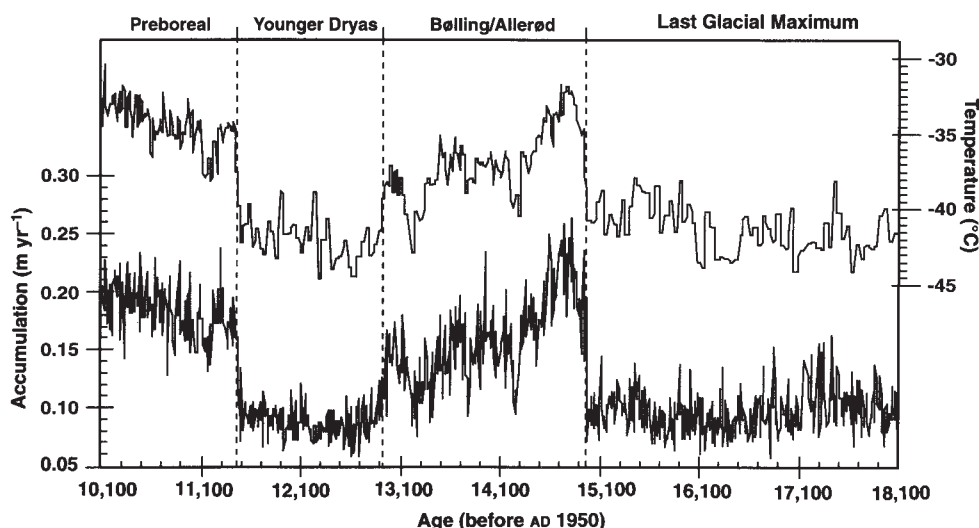
The accumulation and temperature (°C) records in Fig. 1 show strong covariance for major changes between climate states, but less agreement for smaller changes within climate states. If accumulation (b) is controlled thermodynamically by the ability of warmer air to deliver more moisture, we would expect approximately an Arrhenius dependence on absolute temperature: $b \propto \exp[-Q/(RT)]$, where R is the gas constant and Q is the activation energy^{19,20}. Atmospheric models for modern Greenland for the column-averaged moisture content in the

TABLE 1 Sensitivity of snow accumulation rate to temperature change

Time period	Sensitivity* (% per degree K)
Holocene	$\sim 0.9 \pm 0.64$
Preboreal	$\sim 6.8 \pm 0.58$
Younger Dryas	$\sim 1 \pm 0.86$
Bølling/Allerød	$\sim 7.5 \pm 0.52$
Last Glacial Maximum	$\sim 0.2 \pm 0.69$
Entire data set	$\sim 9.5 \pm 0.28$

* 95% confidence limits given.

FIG. 1 Accumulation (bottom trace, 10-yr averages) and $\delta^{18}\text{O}$ (top trace, 1-m averages) against age (years before AD 1950).



troposphere^{2,14} yield $Q_0 \approx 20 \text{ kJ mol}^{-1}$, indicating an increase of $\sim 4\%$ in accumulation per degree warming for the thermodynamic contribution. Models with accumulation from just the lower troposphere¹⁹ yield a larger sensitivity, as much as $\sim 10\%$ per degree. The hypothesis of thermodynamic control on accumulation thus leads us to expect an increase in accumulation of 4% per degree or slightly more, but $<10\%$ per degree.

A windowed (see Fig. 2 legend) cross-correlation of $\ln b$ versus $1/T$ yields Q values shown by colour-coding in Fig. 2. Depositional and diagenetic noise (for example owing to snowdrifting) affects short-period records, but can be neglected over sufficiently long averages^{21–23}. We thus calculate Q from annually resolved data and from running means of 3, 5, 7, ... 151 years, with the window length taken as ten times the running-mean length in each case. We interpret patterns in Fig. 2 that persist across smoothing lengths of decades to centuries as signal rather than noise.

Application of Student's t -test to various smoothing lengths gives high confidence that the apparent activation energy differs from the thermodynamically expected value, Q_0 , at almost all times (ref. 24 and Table 1). Accumulation increases only weakly with temperature (as low as 1% per degree or less) in both cold climates (the Last Glacial Maximum and the Younger Dryas) and warm climates (the Holocene, warm times during the Bølling/Allerød and parts of the Preboreal) (Fig. 3 and Table 1), but increases strongly with temperature (as much as 13% per degree) during changes between warm and cold climates (end of Last Glacial Maximum, inter-Allerød cold periods, onset and end of Younger Dryas, events in the Preboreal; Fig. 2). Allowing for the probable effects of circulation changes on the isotope-temperature relation as discussed above increases the sensitivity at the major climate changes, to perhaps $20\text{--}25\%$ per degree. 'Events' showing a strong sensitivity of accumulation to temperature are most apparent at the Younger Dryas-Bølling/Allerød oscillation, but persist up to a significant event (not shown) at $\sim 8.0\text{--}8.2$ kyr ago. The rest of the Holocene appears to be more uniform.

Our data are broadly consistent with other published records from Greenland. Some shorter Holocene data sets have yielded essentially no correlation between accumulation and temperatures²⁵, whereas others give typically $4\text{--}5\%$ per degree but with little correlation for some sites and smoothing lengths²². The only other long data set is from the GRIP (Greenland Ice-core Project) core²⁶, 30 km east of GISP2. The single regression through the GRIP combined glacial and interglacial data yields a result indistinguishable from ours within probable error limits, including differences in sampling strategies ($\sim 8\%$ per degree GRIP, $\sim 9.5\%$ per degree GISP2). It is in using a windowed cross-correlation, however, that we extract the strong time-vari-

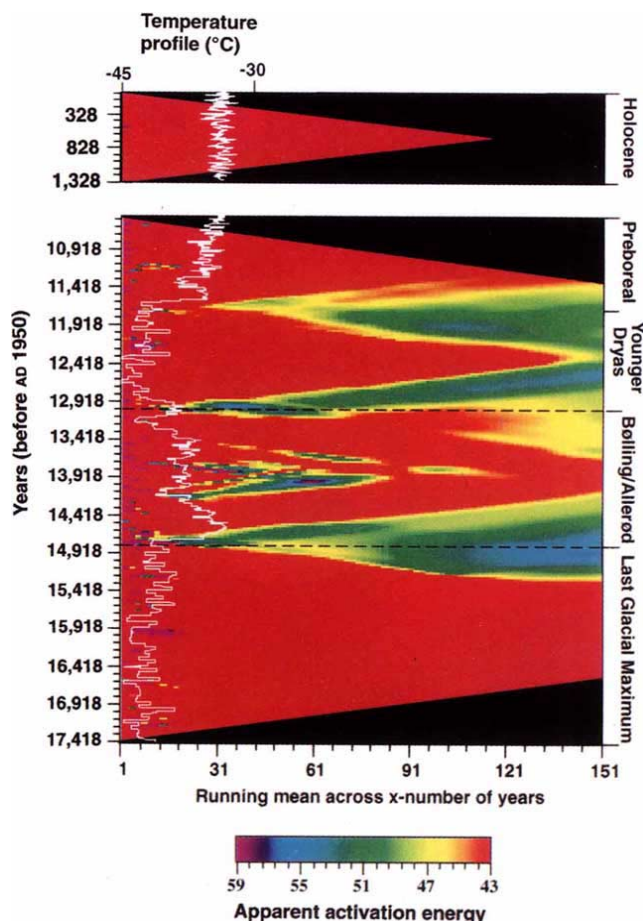
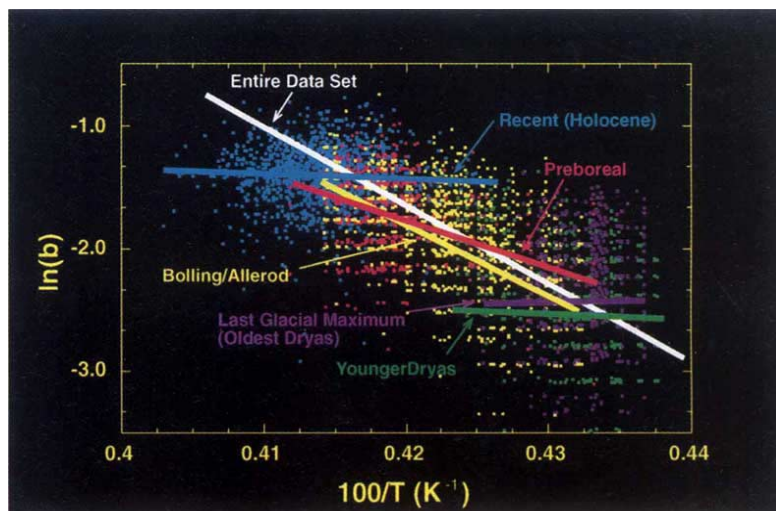


FIG. 2 Apparent activation energy Q in kJ mol^{-1} (see colour bar), determined in each window as the slope of the linear regression of the natural logarithm of accumulation against inverse absolute temperature. Q is plotted with respect to age (years before AD 1950, vertical axis) and smoothing length in years (horizontal axis). Window length is always ten times the smoothing length. For raw data, the regression is performed for years 1–10 (resulting Q plotted at year 5.5), for 2–11 (plotted at year 6.5), and so on through the data set; smoothed data are treated in the same way, but with a longer window. High sensitivity of accumulation to temperature is colour-coded purple, low sensitivity is red. The activation energy for the sublimation of ice is green. This colour-coding optimizes the depiction of the large climate changes. The expected activation energy, Q_0 , is well within the red. The temperature profile calculated from $\delta^{18}\text{O}$, is shown in white.

FIG. 3 Linear regression of the natural logarithm of accumulation b against inverse absolute temperature for each of the major climate events in our data set, and for the entire data set.



tion of the accumulation-temperature relation shown in Fig. 2.

The most direct interpretation of our data is that snow accumulation in central Greenland is controlled primarily by atmospheric dynamics rather than by temperature. In the Holocene, we see weaker dependence of accumulation on temperature than expected for thermodynamic control ($\sim 0.9 \pm 0.6\%$ per degree observed with 95% confidence, compared with the expected 4% per degree or more). Others have also emphasized the importance of atmospheric dynamics relative to temperature in controlling recent accumulation^{2,27-29}. In particular, Bromwich *et al.*²⁷ demonstrated that changes in snow accumulation in central Greenland over the period 1963-88 were controlled primarily by changes in the strength and position of the dominant storm track in the region, with little relation to temperature. Thus simple warming of central Greenland is unlikely to lead to a large increase in snow accumulation³⁰, unless it affects circulation patterns or storminess.

During major climate transitions, accumulation varies more than can be explained thermodynamically (>99% confidence). The greatly reduced snow accumulation during cold conditions then indicates either reduced storminess or changes in storm tracks. During cold periods, however, the steepened equator-to-pole temperature gradient and the greatly increased atmospheric loading of wind-entrained aerosols³¹ argue for stronger storms. We thus infer that there is storm-track motion towards Greenland during major warmings, and away from Greenland during major coolings¹⁷. This is consistent with the prevailing view of

North Atlantic climate^{28,32}, which is that cold periods are characterized by a southward shift and strengthening of the polar front, probably in response to reduced deep-water formation and expanded sea-ice coverage in the North Atlantic. This response would strengthen storms but steer them towards southern Europe rather than north towards Greenland.

The best estimate by the Intergovernmental Panel on Climate Change (IPCC)³³ of sea-level change in a greenhouse-warmed world assumes that accumulation in Greenland will increase about 4% per degree, although with significant uncertainty; from our analysis and other Holocene results^{22,25}, this is about as large a sensitivity as can be expected and a smaller accumulation-rate increase is more probable. For Antarctica, the IPCC cites several lines of evidence for accumulation-rate sensitivity to warming, but relies most heavily on the thermodynamic hypothesis supported by the ice-core record of accumulation-rate and temperature change from glacial to interglacial conditions. Bromwich and Robasky³⁴ showed that accumulation in Antarctica is largely dominated by weather effects rather than the thermodynamically controlled equilibrium water-vapour content of the atmosphere. Our evidence for strong time-dependence of the accumulation-rate/temperature relation in Greenland raises questions about whether the glacial-interglacial signal in Antarctica really supports thermodynamic control of accumulation; the signal may be reflecting changes in storminess. Thus, it may be prudent to plan for somewhat larger future sea-level rises than that of the IPCC 'best estimate' case. □

Received 18 July; accepted 15 November 1994.

- Wigley, T. M. L. & Raper, S. C. B. *Nature* **357**, 293-300 (1992).
- Oerlemans, J. *Holocene* **1**, 40-49 (1991).
- Peel, D. A. *Weather* **46**, 95-102 (1991).
- Alley, R. B. *et al.* *Nature* **362**, 527-529 (1993).
- Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. J. & Jouzel, J. *Nature* **366**, 552-554 (1993).
- Taylor, K. C. *et al.* *J. Glaciol.* **38**, 325-332 (1992).
- Ram, M. & Illing, M. *J. Glaciol.* (in the press).
- Schott, C., Waddington, E. D. & Raymond, C. F. *J. Glaciol.* **38**, 162-168 (1992).
- Taylor, K. C. *et al.* *Nature* **366**, 549-552 (1993).
- Cuffey, K. M., Alley, R. B., Grootes, P. M. & Anandakrishnan, S. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **98**, 265-268 (1992).
- Cuffey, K. M., Alley, R. B., Grootes, P. M., Bolzan, J. M. & Anandakrishnan, S. *J. Glaciol.* **40**, 341-349 (1994).
- Dansgaard, W. *et al.* *Nature* **364**, 218-220 (1993).
- Gundestrup, N. S., Dahl-Jensen, D., Johnsen, S. J. & Rossi, A. *Cold Regions Sci. Technol.* **21**, 399-402 (1993).
- Johnsen, S. J., Dansgaard, W. & White, J. W. C. *Tellus* **41B**, 452-468 (1989).
- Dansgaard, W., White, J. W. C. & Johnsen, S. J. *Nature* **339**, 532-534 (1989).
- Charles, C. D., Rind, D., Jouzel, J., Koster, R. D. & Fairbanks, R. G. *Science* **263**, 508-511 (1994).
- Rind, D., Peteet, D., Broecker, W. S., McIntyre, A. & Ruddiman, W. *Clim. Dyn.* **1**, 3-33 (1986).
- Fisher, D. A. in *The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) 267-293 (NATO ASI Ser. I, Vol. 2, Springer, Berlin, 1992).

- Robin, G. de Q. *Phil. Trans. R. Soc.* **B280**, 143-168 (1977).
- Jouzel, J. *et al.* *Quat. Res.* **31**, 135-150 (1989).
- Fisher, D. A. *et al.* *Nature* **301**, 205-209 (1983).
- Clausen, H. B., Gundestrup, N. S., Johnsen, S. J., Bindshadler, R. & Zwally, H. J. *Ann. Glaciol.* **10**, 10-15 (1988).
- Benoist, J. P., Jouzel, J., Lorius, C., Merlivat, L. & Pourchet, M. *Ann. Glaciol.* **3**, 17-22 (1992).
- Kapsner, W. R. thesis, Pennsylvania State Univ. (1994).
- Reeh, N. *et al.* *J. Glaciol.* **20**, 27-30 (1978).
- Dahl-Jensen, D., Johnsen, S. J., Hammer, C. U., Clausen, H. B. & Jouzel, J. in *Ice in the Climate System* (ed. Peltier, W. R.) 517-532 (NATO ASI Ser. I, Vol. 12, Springer, Berlin, 1993).
- Bromwich, D. H., Robasky, F. M., Keen, R. A. & Bolzan, J. F. *Clim.* **6**, 1253-1268 (1993).
- Serreze, M. C., Box, J. E., Barry, R. G. & Walsh, J. E. *Met. Atmos. Phys.* **51**, 147-164 (1993).
- Fisher, D. A. *Cold Regions Sci. Technol.* **21**, 61-77 (1992).
- Zwally, H. J. *Science* **246**, 1589-1591 (1989).
- Mayewski, P. A. *et al.* *Science* **261**, 195-197 (1993).
- Broecker, W. S. & Denton, G. H. *Geochim. cosmochim. Acta* **53**, 2465-2501 (1989).
- Warrick, R. & Oerlemans, J. in *Climate Change* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 257-281 (Cambridge Univ. Press, 1990).
- Bromwich, D. H. & Robasky, F. M. *Met. Atmos. Phys.* **51**, 259-274 (1993).
- Meese, D. A. *et al.* *Science* (in the press).

ACKNOWLEDGEMENTS. We thank the GISP2 Science Management Office, the Polar Ice Coring Office, the US 109th Air National Guard and GISP2 colleagues for making this research possible. This work was supported by the US NSF, the D. and L. Packard Foundation and NASA-EOS.

Stratigraphic evidence for an early collision between northwest India and Asia

Richard A. Beck*, Douglas W. Burbank*, William J. Sercombe†, Gregory W. Riley†, Jeffrey K. Barndt‡, John R. Berry†, Jamil Afzal‡, Asrar M. Khan‡, Hermann Jürgen‡, Jörgen Metje‡, Amjed Cheema‡, Nasseer A. Shafique‡, Robert D. Lawrence§ & M. Asif Khan||

* Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-0740, USA

† Amoco Production Company, 501 Westlake Boulevard, Houston, Texas 77253, USA

‡ Hydrocarbon Development Institute of Pakistan, Sector 18, H/9, PO Box 1236, Islamabad, Pakistan

§ Department of Earth Sciences, Oregon State University, Corvallis, Oregon 97331-5506, USA

|| National Centre of Excellence in Geology, University of Peshawar, Peshawar 25120, Pakistan

THE collision of India with Asia¹ had a profound influence on late Cretaceous and Cenozoic oceanography², climate³, faunal extinctions⁴ and the motion of at least some of the Earth's lithospheric plates⁵. As the collision ended a period of rapid Indo-Asian convergence⁶, a precise knowledge of its timing (when the crust of the neo-Tethys ocean was completely subducted⁷, at some point along the plate boundary) is important for understanding its wider consequences. But current estimates of the collision age range from 65 to 38 Myr before present^{6,8-11}. Here we report the results of extensive biostratigraphic analyses from Waziristan and Kurram in northwest Pakistan, which show that accretionary-prism and trench strata were first thrust onto the northwest Indian passive

margin after 66 Myr but before 55.5 Myr. After this time, volcanic-arc fragments, the accretionary prism, trench material and imbricates of the north Indian slope were raised to shallow water depths and overlapped by upper Palaeocene shallow-water carbonates and shales¹²⁻¹⁴, indicative of post-collision thrusting in this region. Finally, both the suture and the Indian craton were overlapped by continuous unconformable upper Lower Eocene shallow-marine strata, demonstrating that suturing was complete by 49 Myr.

Before the collision of the northwest Indian craton with Asia, an accretionary prism and an associated composite of previously accreted volcanic arcs and microplates formed the southern margin of Asia above the northward-subducting crust of the neo-Tethys ocean^{7,9}. Due to inevitable irregularities in the pre-collisional margins of India and Asia, possible collisional obliquity and variable rotation of India with respect to Asia during the Cenozoic era, it is likely that collision was not synchronous across the entire collision zone^{6,7}. Nevertheless, the most commonly cited age for early Indo-Asian collision (~50 Myr) comes from Ladakh in the western Himalaya^{7,9,15}.

But interpretation of the Ladakh data is ambiguous for the following reasons: (1) two discrete tectonic events, ophiolite obduction (75–60 Myr) and collision (55–36 Myr), separated by unconformable Palaeocene strata have previously been postulated^{8,9,15}; (2) the actual cause of flexural unconformities¹⁶ is unknown, because flexure of this magnitude could result from intraplate stresses¹⁷ or ophiolite obduction^{7,18} rather than collision itself; (3) Lower and Middle Eocene volcanoclastics of the Indian shelf^{16,19} are not structurally linked to a specific allochthon; (4) the age of unconformable terrestrial strata which overlap both India and Asia in Ladakh is not known with precision (± 13 Myr)⁷; (5) the age of the last marine sediments in the suture zone does not necessarily represent a maximum age for collision^{7,15} as demonstrated here.

We report the results of recent stratigraphic, structural and chronological studies in the Kurram and Waziristan tribal areas

FIG. 1 Main figure, geological map of northwest Pakistan and eastern Afghanistan^{27,58}. Locations: K, Kabul; P, Peshawar. Tectonic boundaries: CF, Chaman fault; GF, Gardez fault; GOFZ, Gwarar Oba fault zone; KNF, Kunar fault; MBT, main boundary thrust; MBZ, mélangé boundary zone; MMT, main mantle thrust; SBF, Sarubi fault; SKF, Safed Koh fault. Lithosomes: CAB, central Afghan block; JIC, Jalalabad igneous complex; KIC, Kohistan igneous complex; KB, Kabul block; KFB, Katawaz Flysch basin (Khojak Flysch); KM, Kahi mélangé; LMUC, Logar mafic-ultramafic complex; NRSB, Nooristan block; WIC, Waziristan igneous complex. Interpretation of the Waziristan igneous complex as part of a volcanic arc²⁶ is bolstered by a suite of intrusive and extrusive rocks that resemble the igneous complexes of the Kohistan arc^{59,60}, including ultramafic rocks with high-pressure amphibole and pyroxene mineral chemistries²⁶. The WIC includes dacite, rhyodacite, andesite, tuff, agglomerate, volcanic breccia and copper deposits^{14,26,61} typical of arc volcanism⁶¹. Thick sections of pyroclastic breccia with decimetre-sized bombs and layers of tuff up to 3 m thick in the upper part of the WIC¹⁴ preclude a mid-oceanic-ridge origin⁶¹. Interbedded Albian carbonate with *Flabellamina* sp., radiolarite and

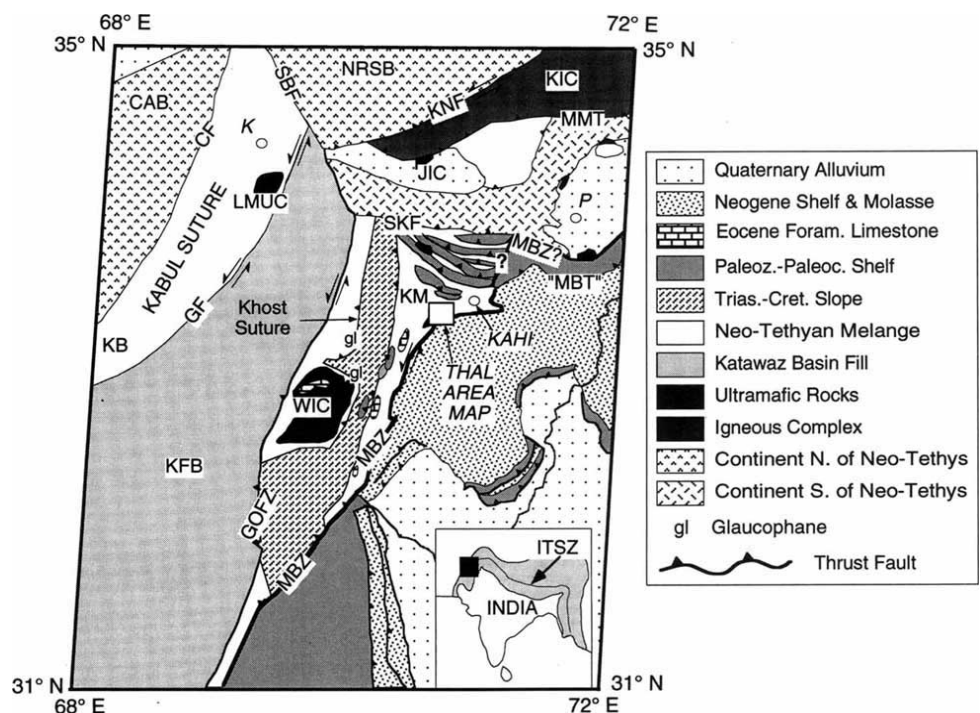
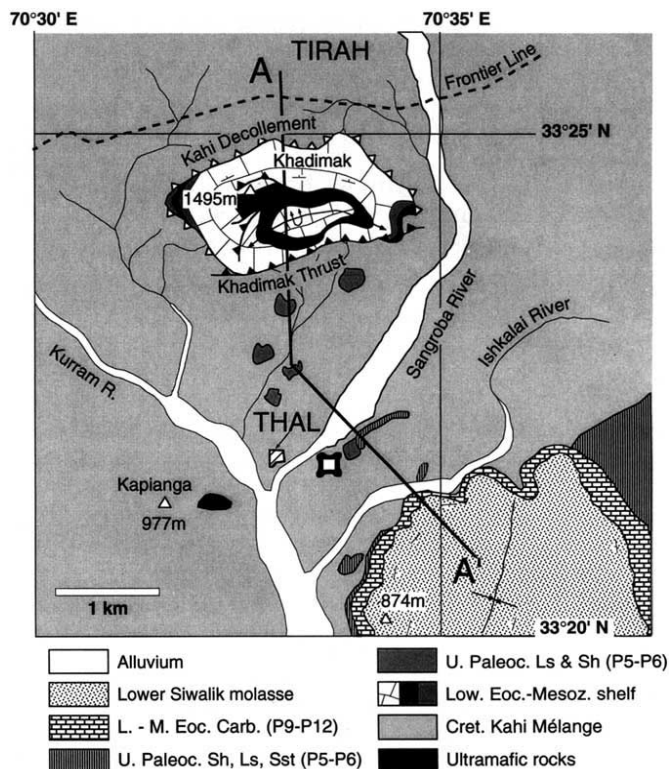


FIG. 2 Geological sketch map of the Thal area¹². The Kahi mélangé contains the Kapianga mafic-ultramafic complex. Together they are composed of dunite, peridotite, pillow basalt^{26,62}, glaucophane-bearing^{14,62} tuff which locally encases Cenomanian *Rotalipora cushmani*, manganese nodules, kilometre-scale blocks of Upper Cretaceous radiolarite²² including *Cryptamphorella conara* and *Amphipyndax pseudoconulus*, Mesozoic belemnite-bearing calcilutite, and Turonian (*Helvetoglobotruncana helvetica*, *Præoglobotruncana stephani*, *Whiteinella brittonensis*, *Marginotruncana* sp.) to Maastrichtian (*Heterohelix navarroensis*, *Globotruncana arca*, *Globotruncana bulloides*) deep-marine shale which contains centimetre-to-metre sized olistoliths including coral fragments, turbidites containing volcanoclastic sandstone with Campanian–Maastrichtian open marine foraminifera (*Globotruncana gansseri*, *Abathomphalus mayaroensis*) and non-volcanoclastic sandstone²⁰. The Kahi mélangé is a previously unrecognized segment of the neo-Tethyan accretionary prism/trench complex which crops out over >15,000 km² of the tribal areas of northwest Pakistan and eastern Afghanistan²⁰ (Figs 1 and 2). This rock assemblage represents a link between segments of the neo-Tethyan suture on the northern and western margins of the Indian craton^{63,64} (Fig. 1), and serves to explain the abundance of published occurrences of ultramafic rocks between Waziristan–Kurram and well accepted Indus suture outcrops north and west of the Peshawar basin in northern Pakistan^{21,60,65–67} (Fig. 1). The youngest strata preserved in the accretionary prism/trench complex date from late to latest Maastrichtian and provide a maximum age constraint for the beginning of collision in this area (Fig. 2). Lithologies: Sh, shale; ls, limestone; Sst, sandstone; Carb, carbonates in general.

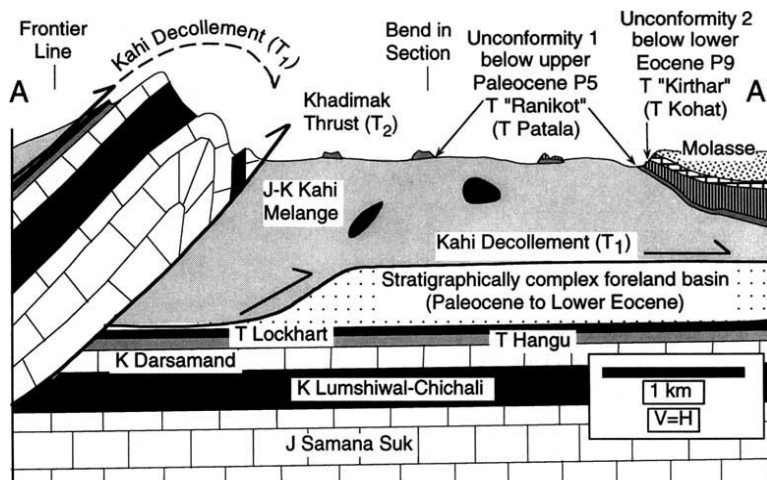


of northwest Pakistan, where distinct episodes of collision and post-collisional thrusting can be discerned in the Palaeocene-to-Eocene stratigraphic record. During the first episode, continental collision, the accretionary prism and trench complex were thrust on to the north-facing continental rise and slope of the Indian craton and raised to shallow water depths before being overlapped by shallow marine Upper Palaeocene strata. During the second episode, volcanic-arc fragments, the accretionary prism, trench complex and imbricates of the north Indian slope and shelf were thrust further across the Indian shelf, largely eroded, and subsequently overlapped by upper Lower Eocene limestones

which are continuous with the Indian shelf sequence. These overlapping and cross-cutting relationships provide the oldest unambiguous constraints on the timing of collision and suturing in this region as well as the entire Himalaya.

Our field work²⁰ and re-interpretation of published data^{12,21,22} indicates that remnants of the Indo-Asian suture are present 100–300 km southeast of its previously recognized trace^{23–25} (Fig. 1). Three major lithologic elements define the suture in Waziristan and Kurram (Fig. 1): (1) the large ultramafic-to-acidic Waziristan igneous complex (WIC)²⁶ which represents a segment of the trans-himalayan arc²⁰ (Fig. 1); (2) the intensely

FIG. 3 Simplified geological cross-section of the Thal area^{12,22} along A–A' in Fig. 2. Unconformity 1 occurs where Upper Palaeocene shelf limestones and shales depositionally overlap the Kahi mélangé¹². Although this contact has been interpreted as a southeast-vergent thrust fault²², our mapping results and those of Davies^{12,21} indicate that it is primarily a depositional contact. The Late Palaeocene age is supported by four species of Palaeocene molluscs¹², four species of Palaeocene corals¹², and new foraminiferal biostratigraphic studies near Kahi which indicate that these strata encompass faunal zones P5 and P6. Unconformity 1 defines initial collision between the Waziristan–Kurram segment of the trans-Himalayan arc and the Indian craton in this region. Initial thrusting of the accretionary prism on to the Indian slope occurred between 66 and 55.5 Myr. This scenario is corroborated by the development of an adjacent marine to non-marine foreland basin on the northwest Indian craton⁶⁸ as well as significant pre-Eocene deformation of the shelf sequence in northern Pakistan⁶⁹. Unconformity 2 developed when the arc/prism/trench complex and Indian slope were thrust farther on to the Indian shelf. Detritus included a massive influx of Jurassic dinoflagellates and pollen which was redeposited in the adjacent foreland basin. That basin progressed from shale-dominated, marine deposition to subaerial red beds analogous to the Balakot–Murree formation found in the Jhelum re-entrant of northern Pakistan⁷⁰. The



eroded arc/prism/trench complex and the gently folded, unconformable Upper Palaeocene strata were subsequently overlapped above a clear unconformity by shallow-marine strata of the P9 foraminiferal zone (50–49 Myr), substantially after collision. V = H, vertical = horizontal. Geological periods, J, Jurassic; K, Cretaceous; T, Tertiary. The names that follow these abbreviations refer to layers or masses of rock.

deformed ultramafic- and glaucophane-bearing Kahi mélange, which represents the neo-tethyan accretionary prism and trench complex (Figs 1 and 2); and (3) non-volcaniclastic, deep-to-shallow marine carbonates, shales and sandstones of the northwest Indian shelf and slope^{14,27,28}.

The WIC and Kahi mélange must be placed in the context of Afghan tectonics to address the issue of Indo-Asian collision because two sutures separate the northwest Indian craton and central Afghanistan (Fig. 1)²⁷. The central Afghan block (CAB) has been a part of Asia since at least the early Cretaceous period²⁷. The northwestern suture (the Kabul suture), closed during the Palaeocene²⁷ when the Gondwanan Kabul block was accreted to the southeastern margin of the CAB extinguishing the northern Kandahar arc (111–69 Myr)²⁹. Unconformable Early Eocene strata overlie the Kandahar arc, imbricates of ultramafic rock, metamorphosed Gondwanan shelf and slope of the Kabul block, and remnants of a Late Cretaceous accretionary prism/trench complex^{14,28,30–41}.

Igneous rocks of the WIC were thrust onto the northwest Indian craton during the Palaeocene^{14,27}. Re-interpretation of the WIC as a volcanic arc or arc fragment²⁶, and of strata along the southeastern boundary of Waziristan and Kurram as imbricates of Indian slope and accretionary prism/trench complex²⁰, implies that overlying unconformable Upper Palaeocene (P5) and Upper Eocene (P9) strata^{12,13,42–46} date collision and full suturing rather than an intra-oceanic ophiolite obduction event followed by collision. Moreover, the similar Palaeozoic to Eocene tectonostratigraphies of the Kabul (northwest) and Khost–Waziristan (southeast) Sutures²⁷ suggest they are the same suture, now repeated in cross-section and map view (Fig. 1). Therefore our revision of the geology of Waziristan and Kurram supports the hypothesis that the Kabul block (Fig. 1) represents a corner of the northwest Indian craton which was broken off and extruded⁴⁷ towards the southwest⁴⁸ after the collision of northwest India and Asia.

The newly recognized mélange boundary zone (MBZ) separates the Kahi mélange from regions of exclusively Indian shelf strata. In the northern and southern parts of the study area (Fig. 1), the MBZ is generally a southeast-vergent thrust above the Permo-Triassic to Cretaceous shelf sequence. In its central part, however, the MBZ is usually represented by one or more Tertiary unconformities. The two oldest unconformities are defined by the overlap of parts of the Kahi mélange by parautochthonous strata of the Indian shelf^{12,21,22} (Figs 2 and 3). Despite local structural disruption due to exploitation of some of the unconformities by back thrusts, the fundamental stratigraphic relationships of the unconformities are locally preserved near the town of Thal (Figs 1 and 2). The unconformable Palaeocene strata which overlie the prism/trench/slope complex indicate they were juxtaposed before the end of the P5 foraminiferal zone (>55.5 Myr)^{12,13,21,42–46,49–53} (Figs 2 and 3).

During continued thrusting, the Kahi mélange overran Lower Palaeocene shelf limestones and parts of its own foreland basin (Fig. 3). This thrust contact is exposed on the west, north and east sides of Khadimak mountain (Fig. 2), where a regional décollement places the Kahi mélange above Late Palaeocene to Early Eocene (P6) Indian shelf strata (Fig. 3). North of Thal, numerous thrust imbricates of shelf strata cut upward through the basal décollement of the Kahi mélange in a style analogous to that found at Khadimak mountain (Fig. 2)²². Although we can not precisely quantify the amount of shortening represented by transpressional deformation north of the MBZ⁵⁴, the present locations of these imbricates with respect to the MBZ suggest a lower limit of 180 km for the amount of shortening during thrusting of the Kahi mélange on to the northwest Indian craton.

Upper Lower Eocene (P9) strata at Thal fort (Fig. 2) can be traced from their unconformable contact with the supra-mélange Palaeocene strata along strike to a continuous stratal succession above the north Indian craton^{22,55}. This is the first continuous pre-Middle Eocene physical link between a neo-Tethyan accre-

tionary prism/trench complex and the Indian craton ever reported. Because the Kabul suture (whatever its relationship to the Khost–Waziristan suture) had already closed during the Palaeocene^{14,27}, the upper Lower Eocene shallow-marine strata at Thal and elsewhere in Waziristan^{12,13,21,42–46,49–53} also date full suturing of the northwest Indian craton with Asia. The similarity of the timing of full suturing of the WIC/Kahi mélange and of oceanic igneous complexes farther south along the western Indian plate margin⁵⁶ suggests simultaneous modification (>49 Myr) of the northern convergent and western transform plate boundaries. This is consistent with the idea of post-collisional transpressional transitions from convergent to transform plate margins^{54,57}. Because the suture has been “stitched” to the Indian craton since 49 Myr, and because <100 km of shortening has occurred south of the suture since then⁵⁴, nearly all Indo-Asian convergence that occurred after 49 Myr ($\geq 2,000$ km)⁶ has been accommodated north of the suture in this area. □

Received 9 May; accepted 24 November 1994.

- Argand, E. *La Tectonique de l'Asie* 171–372 (C.R. Int. Geol. Congress, Brussels, 1924).
- Stille, P. *Geology* **20**, 387–390 (1992).
- Kerrick, D. & Caldeira, K. *Chem. Geol.* **108**, 201–230 (1993).
- Jaeger, J. J., Courtillot, V. & Tapponnier, P. *Geology* **17**, 316–319 (1989).
- Patriat, P. & Achache, J. *Nature* **311**, 615–621 (1984).
- Dewey, J. F., Cande, S. & Pitman, W. C. *Eclogae Geol. Helv.* **82**, 717–734 (1989).
- Searle, M. P. et al. *Geol. Soc. Am. Bull.* **98**, 678–701 (1987).
- Searle, M. P. *J. struct. Geol.* **8**, 923–936 (1986).
- Searle, M. P., Cooper, D. W. J. & Rex, A. J. *Phil. Trans. R. Soc. A* **A326**, 117–150 (1988).
- Klootwijk, C. T. in *Structural Geology of the Himalaya*, Delhi (ed. Saklani, P. S.) 307–360 (Hindustan Press, Delhi, 1979).
- Klootwijk, C. T., Gee, J. S., Peirce, J. W., Smith, G. W. & McFadden, P. L. *Geology* **20**, 395–398 (1992).
- Davies, L. M. *J. Geol. Soc. Lond.* **83**, 260–290 (1926).
- Davies, L. M. *Nature* **139**, 414–415 (1937).
- Cassaigneau, C. thesis, Univ. Sci. Tech., Languedoc (1979).
- Searle, M. P., Cooper, D. W. J. & Pickering, K. T. *Tectonophysics* **174**, 301–314 (1990).
- Garzanti, E., Baud, A. & Mascle, G. *Geodinam. Acta* **1**, 297–312 (1987).
- Cloetingh, S. *Geology* **14**, 617–620 (1986).
- Searle, M. P., Kelemen, P. B., Reuber, I. & Fuchs, G. *J. struct. Geol.* **8**, 129–132 (1988).
- Gaetani, M. & Garzanti, E. *Am. Ass. Petrol. Geol.* **75**, 1427–1446 (1991).
- Beck, R. A., Sercombe, W. J. & Burbank, D. W. *Geol. Soc. Am. Abstr. Progr.* **25**, 122 (1993).
- Davies, L. M. *Mem. geol. Surv. India, Paleont. Indica (new ser.)* **15**, 1–225 (1930).
- Meissner, C. R., Hussain, M., Rashid, M. A. & Sethi, U. B. *Prof. Pap. U.S. geol. Surv.* **716**, 1–24 (1975).
- Carbonnel, J. P. *Mem. Soc. géol. Fr.* **8**, 145–152 (1977).
- Treloar, P. J. & Izatt, C. N. in *Himalayan Tectonics* (eds Treloar, P. J. & Searle, M. P.) 69–87 (Geol. Soc. London, 1993).
- Brookfield, M. E. *Sedim. Geol.* **84**, 1–35 (1993).
- Ahmad, W. & Hamidullah, S. *Geol. Bull. Univ. Peshawar* **20**, 161–180 (1987).
- Tapponnier, P., Mattauer, M., Proust, F. & Cassaigneau, C. *Earth planet. Sci. Lett.* **52**, 355–371 (1981).
- Mennessier, G. *Bull. Soc. géol. Fr.* **24**, 321–330 (1982).
- Wolffart, R. & Wittekindt, H. *Geologie von Afghanistan* (Gebrüder Borntrager, Berlin, 1980).
- Griesbach, C. L. *Rec. geol. Surv. India* **25**, 59–109 (1982).
- Cizancourt, H. *Mem. Soc. géol. Fr.* **39**, 5–28 (1938).
- Kaever, M. C. *Soc. géol. Fr.* **7**, 188–189 (1961).
- Kaever, M. *Geol. Jb.* **82**, 131–142 (1964).
- Tapponnier, P., Mattauer, M., Proust, F. & Cassaigneau, C. *Earth Planet. Sci. Lett.* **52**, 355–371 (1981).
- Mennessier, G. & Kaever, M. C. *Soc. géol. Fr. D282*, 1929–1932 (1976).
- Mennessier, G. *C. r. hebd. Séanc. Acad. Sci., Paris D282*, 1581–1583 (1976).
- Mennessier, G. *C. r. hebd. Séanc. Acad. Sci., Paris D270*, 1427–1429 (1970).
- Mennessier, G. *C. r. Soc. géol. Fr.* **271**, 64–66 (1970).
- Mennessier, G. *C. r. Soc. géol. Fr.* **269**, 314–316 (1969).
- Mennessier, G. *C. r. hebd. Séanc. Acad. Sci., Paris D270*, 2521 (1970).
- Mennessier, G. *Mem. Soc. géol. Fr.* **8**, 153–168 (1977).
- Davies, L. M. *Nature* **142**, 296 (1938).
- Davies, L. M. in *Proc. 6th Pacific Sci. Congr.* (ed. Lipman, C. B.) Part 2, 483–501 (Univ. California Press, Berkeley, 1940).
- Davies, L. M. *Int. Geol. Congr., Rep.* 18 73–76 (Proc. Int. Paleont. Union, London, 1950).
- Davies, L. M. *C. r. Soc. géol. Fr.* **2**, 22–23 (1938).
- Davies, L. M. *Nature* **141**, 202 (1938).
- Molnar, P. & Tapponnier, P. *Science* **189**, 419–426 (1975).
- Fourcade, E. et al. *Bull. Soc. géol. Fr.* **162**, 1197–1208 (1993).
- Stuart, M. *Rec. geol. Surv. India* **54**, 87–102 (1922).
- Smith, F. H. *Rec. geol. Surv. India* **28**, 106–110 (1895).
- Davies, L. M. *Nature* **135**, 188 (1935).
- Davies, L. M. *J. geol. Soc. Lond.* **46**, 199–231 (1939).
- Davies, L. M. *J. geol. Soc. Lond.* **99**, 63–79 (1942).
- Pivnik, D. A. & Sercombe, W. J. in *Himalayan Tectonics* (eds Treloar, P. J. & Searle, M. P.) 559–580 (Geol. Soc. London, 1993).
- Meissner, C. R., Master, J. M., Rashid, M. A. & Hussain, M. *Stratigraphy of the Kohat Quadrangle, Pakistan* (US Geol. Surv., Washington, DC, 1974).
- Allemann, F. in *Geodynamics of Pakistan* (eds Farah, A. & DeJong, K. A.) 215–242 (Geol. Surv. Pakistan, Quetta, 1979).
- Woodcock, N. H. *Phil. Trans. R. Soc. Lond.* **A317**, 13–29 (1986).
- Lawrence, R. D., Kahn, S. H. & Nataka, T. *White Paper* 1–35 (Geol. Surv. Pakistan, Quetta, 1991).
- Jan, M. Q., Khan, A. & Windley, B. F. *Econ. Geol.* **80**, 294–306 (1983).

60. Khan, M. A., Jan, M. W., Windley, B. F., Tarney, J. & Thirwall, M. F. in *Tectonics of the Western Himalayas* (eds Malinconico, L. L. & Lillie, R. J.) 75–94 (Geol. Soc. Am., Boulder, 1989).
61. Hyndman, D. W. *Petrology of Igneous and Metamorphic Rocks* (McGraw-Hill, New York, 1985).
62. Hayden, H. H. *Rec. geol. Surv. India* **28**, 109–117 (1896).
63. Sengor, A. M. *The Cimmeride Orogenic System and the Tectonics of Eurasia 1–82* (Geol. Soc. Am., Boulder, 1984).
64. Sengor, C. A. M. *Rev. Earth planet. Sci.* **15**, 213–244 (1987).
65. Tahirkehi, R. A. K., Mattauer, M., Proust, F. & Tapponnier, P. in *Geodynamics of Pakistan* (eds Farah, A. & DeJong, K. A.) 125–130 (Geol. Surv. Pakistan, Quetta, 1979).
66. Jan, M. Q. *Geol. Bull. Univ. Peshawar* **7–8**, 73–78 (1975).
67. Coulson, A. L. *Rec. geol. Surv. India* **75**, 1–55 (1940).
68. Wells, N. A. thesis, Univ. Michigan (1984).
69. Yeats, R. S. & Hussain, A. *Geol. Soc. Am. Bull.* **99**, 161–176 (1987).
70. Bossart, P. & Ottiger, R. *Eclogae geol. Helv.* **82**, 133–165 (1989).

ACKNOWLEDGEMENTS. We thank S. H. Afridi, Y. Mahsud, A. Z. Khan, the South Waziristan, Tochi, Thal, and Kurram Scouts and the Shawal Rifles for providing logistical support, Reservoirs Inc. for performing sandstone and D. Mayo for supervising the XRF analyses. We also thank I. Hussain, I. Rahman, M. Shahid and M. Alam for providing assistance in the field and office, and G. A. Davis, A. Meigs, J. Verges, S. Friedmann, C. Carlson, M. Chichanski and P. Molnar for reviews. The Department of Earth, Atmospheric and Planetary Sciences at MIT provided facilities for the preparation of this manuscript. This work was supported by Amoco; D.W.B. thanks the US NSF and the National Geographic Society for grants, and R.A.B. acknowledges Fulbright fellowships, and scholarships from the Achievement Rewards for College Scientists.

An experimental replication of upper-mantle metasomatism

N. W. A. Odling

Department of Geology and Geophysics, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JW, UK

INTERACTION of melts ascending from the asthenosphere with the lithospheric mantle through which they pass is widely accepted to be a significant process by which the lithosphere can acquire distinctive trace-element and isotopic signatures^{1–3}. Despite the importance of this process in the genesis of the lithosphere, the nature of these 'metasomatic' reactions is still debated^{2–5}. Here I report the results of an experiment designed to investigate the mechanism by which such metasomatism could occur. A natural sample of Group I kimberlite, a low-volume partial melt of asthenospheric origin, is reacted with a refractory peridotite (harzburgite) in a thermal gradient at high pressure to simulate the interaction between a dyke and its host rock in the lower lithosphere. The harzburgite develops several contrasting mineralogical zones, the coolest of which contains olivine, orthopyroxene, clinopyroxene, phlogopite, amphibole, Nb–Cr-rutile and minor zircon. The formation of mica and amphibole is accompanied by increases in trace-element concentrations and Rb/Sr, Rb/Ba and Pb/U ratios—features characteristic of metasomatized xenoliths recovered from southern African kimberlites⁵. These results support the suggestion, based on temporal association and isotopic data, that at least some mantle metasomatism is related to the interaction of Group I kimberlite with refractory peridotite in the lower lithosphere⁶.

The mantle xenoliths that occur in kimberlites have received much attention (for example, refs 7–10) because they provide the best direct evidence of the nature of the lithospheric mantle. Among the xenolith types that occur in the Group I kimberlites of southern Africa (see Fig. 1 legend for definition of Group I kimberlite), some show evidence of chemical alteration (metasomatism) in the mantle^{11,12}. In addition to the usual olivine, pyroxenes and garnets typical of many peridotites, these 'metasomites' contain distinctive phlogopite ± alkali-rich amphibole and, in some cases, Nb–Cr-rutile and Ba-titanates. Trace-element studies have shown that the metasomatism has resulted in the development of high Rb/Sr, Rb/Ba and K/Ba, and an enrichment in light rare-earth elements. The metasomites also have high ⁸⁷Sr/⁸⁶Sr (ϵ_{Sr} from +18 to +80; ϵ notation defined in Fig. 2 legend) at ¹⁴³Nd/¹⁴⁴Nd only slightly lower than that of the bulk Earth (ϵ_{Nd} from 0 to –7)¹. Metasomatism is thus an

important process by which the lower lithosphere becomes enriched in both trace and major elements. The nature of the metasomatic 'fluid' (which could be an aqueous fluid or a melt) remains debated, however, because there is no clear relationship between the metasomite trace-element characteristics and any magma type of deep origin^{1,13}. This has led to the suggestion that metasomatic assemblages are the results of complex open-system reactions in which material is both added from, and removed by, melts and fluids passing through peridotite¹.

To test whether the metasomite xenolith assemblages are the products of interaction between a low-volume melt and refractory lithosphere, I have devised an experiment in which natural, aphanitic Group I kimberlite can interact with refractory peridotite in conditions that simulate reaction between a dyke and its wall-rock. A 10-mm-long capsule is loaded with a 'sandwich' consisting of a layer of natural aphanitic kimberlite (S30; ref. 14) overlying a layer of synthetic refractory peridotite (LBM10; ref. 15) in the proportions 1:9 by weight. This capsule is enclosed in an outer capsule containing magnesite + periclase + graphite + H₂O to buffer the oxygen fugacity of the experiment close to the enstatite + magnesite + olivine + graphite (EMOG) equilibrium¹⁶. The capsule assembly is then placed in a standard piston–cylinder high-pressure apparatus in such a way that at run conditions (30 kbar, 24 h) the kimberlite occupies the furnace hot-spot (1,225 °C), whereas the refractory peridotite extends downwards into the thermal gradient to reach a temperature of 1,000 °C at the base of the capsule. The conditions at the capsule base thus simulate those found at ~90 km depth in a geotherm of 40 mW m^{–2}. After quenching, the capsule released 2 wt% of a water-rich fluid, indicating vapour-saturation at run conditions.

The run products consist of six transverse layers of contrasting colour and texture. The highest-temperature layer consists of a small volume (~0.5 mm thick) of radiating, bladed crystals interpreted to be the quench products of residual kimberlite melt. Below are a series of peridotite types of decreasing grain size. The two higher-temperature peridotite layers consist of 2 mm of Cr-spinel + olivine (Cr-sp + ol) overlying 2 mm of Cr-sp + ol + clinopyroxene (cpx), both assemblages containing interstitial glass and phlogopite (phl) crystals. Below these assemblages is a sequence of porous peridotites containing no material identifiable as quench products after a melt phase. This sequence runs

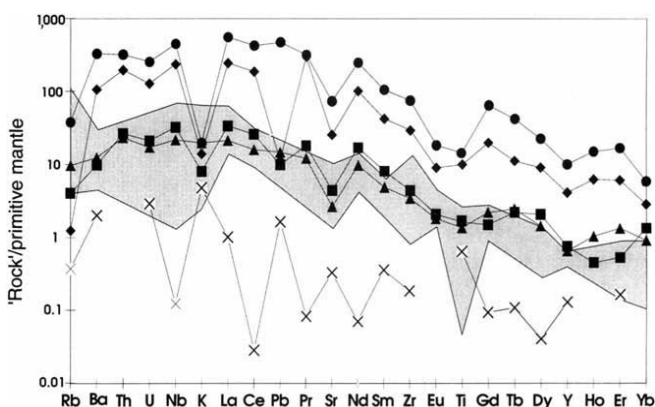


FIG. 1 Bulk trace-element patterns determined by secondary-ion mass spectrometry of fused samples using the method described in ref. 24. Starting materials: circles, aphanitic kimberlite S30; crosses, synthetic harzburgite LBM10. Run products: diamonds, residual kimberlite melt; squares, phlogopite lherzolite; triangles, phlogopite, K-rich lherzolite. Shaded area: maximum and minimum of range shown by phlogopite, K-rich lherzolite peridotite xenoliths⁴. Kimberlites are volcanic rocks that are SiO₂-poor and relatively rich in MgO and K₂O; Group I kimberlites are poorly micaceous and have relatively low abundances of SiO₂, MgO and K₂O compared with the mica-rich Group II kimberlites.

TABLE 1 Experimental starting mixes, run products and natural metasomite compositions

	1	2	3	4	5	5a	6	6a	7	8	9	10
SiO ₂	25.0	45.0	45.1	47.3	57.0		44.6		0.2	0.1	32.5	32.9
TiO ₂	2.69	0.05	0.36	0.33	1.02		1.51		1.07	90.60	—	0.08
Al ₂ O ₃	4.38	1.25	0.80	1.82	1.48	ND-1.79	10.95	7.99-11.98	5.60	0.16	0.20	—
Cr ₂ O ₃	0.41	0.29	0.37	0.14	0.18	1.18-0.21	0.21	ND-0.54	68.32	3.37	0.04	0.02
FeO (total)†	9.42	6.57	3.20	4.90	1.33	3.45-2.30	2.30	2.12-5.31	6.35	0.60	0.26	0.14
NiO	ND	0.30	0.18	0.23	0.07		0.20		0.07	ND	ND	—
MnO	0.24	0.10	0.13	0.12	ND		0.07		0.35	ND	ND	—
MgO	24.76	42.59	44.8	42.5	22.28		26.54		17.72	—	0.40	—
CaO	15.63	0.82	4.60	1.74	8.08		ND		0.02	—	—	0.08
K ₂ O	1.08	0.14	0.24	0.60	4.86		10.34		0.03	—	—	—
Na ₂ O	0.87	0.16	0.21	0.16	2.86	3.25-4.72	0.17	ND-0.78	—	—	—	—
Nb ₂ O ₅	382*	0.11*	21*	18*	—	—	—	—	—	2.59	—	—
ZrO ₂	980*	2.4*	45*	45*	—	—	—	—	—	1.39	64.08	65.8
HfO ₂	—	—	—	—	—	—	—	—	—	—	1.31	1.32
Rb*	25	0.2	4	6	—	—	—	—	—	—	—	—
Sr*	1548	6.7	60	56	—	—	—	—	—	—	—	—
Th*	27	0.02	1.9	2.0	—	—	—	—	—	—	—	—
U*	5.4	0.06	0.35	0.36	—	—	—	—	—	—	—	—
Pb*	31	0.02	0.07	0.11	—	—	—	—	—	—	—	—
H ₂ O	6.47	1.82	—	—	—	—	—	—	—	—	—	—
CO ₂	5.87	0.11	—	—	—	—	—	—	—	—	—	—
Total	99.99	98.07	100.00	100.00	96.92		99.73		99.67	99.38	98.9	100.34

Analyses were obtained by electron probe microanalysis (major elements) and secondary-ion mass spectrometry (trace elements) of samples fused using the method described in ref. 24. **1**, Natural aphanitic kimberlite S30. **2**, Refractory harzburgite LBM10. **3**, Experimental phlogopite lherzolite metasomite (normalized to 100%). **4**, Experimental phlogopite, K-richterite metasomite (normalized to 100%). **5**, Amphibole from experimental phlogopite, K-richterite metasomite. **5a**, Range in composition reported for amphiboles in phlogopite, K-richterite metasomites from southern Africa (ref. 4). **6**, Phlogopite from experimental phlogopite, K-richterite metasomite. **6a**, Range in composition reported for phlogopites in phlogopite, K-richterite metasomites from southern Africa (ref. 4). **7**, Spinel from experimental spinel dunite. **8**, Nb,Cr-rutile from experimental phlogopite, K-richterite metasomite. **9**, Zircon from experimental phlogopite, K-richterite metasomite. **10**, Zircon from mantle xenolith BD3024 (J. Dawson, unpublished data).

* Quoted as p.p.m.; —, not determined; ND, not detected; † total iron expressed as FeO.

from ol + cpx + phl (phlogopite wehrlite, 1.2 mm) through ol + orthopyroxene + cpx + phl (phlogopite lherzolite, 2.2 mm) to ol + orthopyroxene + cpx + phl + amphibole with minor Nb-Cr-rutile and zircon (phlogopite, K-richterite lherzolite, 0.2 mm). Representative analysis of these phases are given in Table 1.

The amphibole at the low-temperature end of the capsule is a K-richterite which is similar to those found in metasomite xenoliths (Table 1) but contrasts strongly with the pargasitic amphibole reported in experiments on the pyrolite-H₂O system¹⁷. The experimental richterite falls within the range of metasomite richterites described by Erlank *et al.*⁵, except for being slightly richer in CaO. The rutile, containing 2.6 wt% Nb₂O₅ and 3.4 wt% Cr₂O₃, is present as a minor phase and, like the amphibole, it is similar to that found in metasomite xenoliths. The zircon contains 1.3 wt% HfO₂ and is similar to a zircon reported in a mica + amphibole + cpx metasomatic vein in a peridotite xenolith from Kimberley, South Africa (J. B. Dawson, personal communication).

Bulk trace-element compositions of these assemblages have been determined by secondary-ion mass spectrometry. Figure 1 shows trace-element patterns, normalized to primitive mantle¹⁸, for the starting mixes, run-product assemblages from the two lowest-temperature zones, and natural metasomite xenoliths. The synthetic peridotite starting composition has a highly irregular pattern but is characterized by low overall abundances of trace elements. In contrast, the kimberlite starting composition has high trace-element abundances, with enrichments in large-ion lithophile and light rare-earth elements, and relative depletions in potassium and strontium, characteristic of Group I kimberlites¹⁹. The two experimental metasomite assemblages show significant abundances of all analysed trace elements. Both metasomite patterns are similar and have a profile similar to the kimberlite but at a concentration level ~20 times lower. The patterns show significant enrichments in rubidium and lead in both phlogopite-bearing assemblages. Thus, the metasomatism

has resulted in increases in Rb/Sr, Pb/U, Pb/Th and Sm/Nd, corresponding to characteristics of metasomatizing kimberlite.

A number of points are evident from the textural, mineralogical and compositional characteristics of the metasomites. First, the wehrlite and lherzolite assemblages are porous but contain no material identifiable as products of quenching of a melt phase. The assemblages are, therefore, products of reactions between the H₂O-rich fluid phase and the harzburgite. Second, whereas garnet is present in the mode of the harzburgite LBM10, it is absent in the experimental assemblages; the metasomatism of garnet peridotite by fluids derived from kimberlite has resulted in elimination of the garnet in favour of phlogopite ± K-richterite, in agreement with previous petrographic studies²⁰. Third, the region of most intense metasomatism is furthest from the kimberlite, implying that the intensity of metasomatism was controlled more by the thermal gradient than by the relative mobility of components.

The changes in Rb/Sr, Sm/Nd and Pb/U in the harzburgite resulting from the experimental metasomatism can be used to assess whether the isotopic signatures that would develop over time are comparable to those of the natural metasomite xenoliths from southern Africa^{1,21}. Here, I assume that lithosphere has a Nd-Sr-Pb isotopic signature within the range shown by 150-Myr-old garnet peridotite xenoliths—just before the main phase of Group I kimberlite magmatism in southern Africa¹. Components in $\epsilon_{\text{Sr}}-\epsilon_{\text{Nd}}$ space are shown in Fig. 2. As the metasomatism introduces both Sr and Nd, the initial effect will be a mixing event. As the budgets for Sr and Nd are dominated by the component introduced from the kimberlite, the metasomite will plot close to the kimberlite position in $\epsilon_{\text{Sr}}-\epsilon_{\text{Nd}}$ space immediately after the metasomatic event (component a in Fig. 2). With time, however, the metasomite Sr and Nd isotopes will evolve as a result of the new values of Rb/Sr and Sm/Nd developed during metasomatism. The increase in Rb/Sr in the metasomite assemblages (Table 1) will lead to an increase in ⁸⁷Sr/⁸⁶Sr. In contrast, the Nd isotopic signature will remain relatively unaffected, as

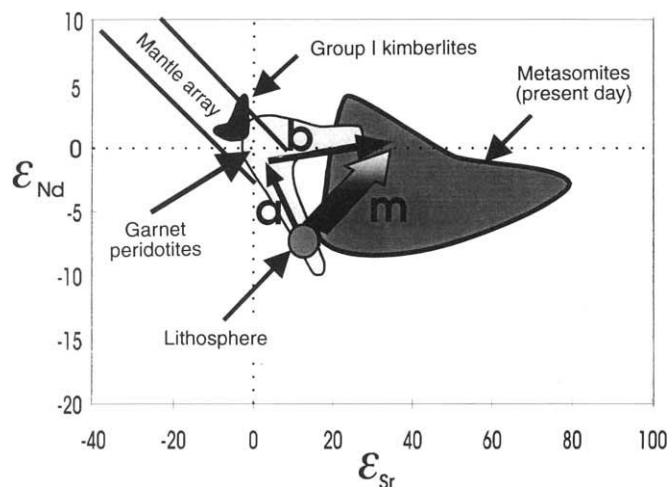


FIG. 2 Schematic $\epsilon_{\text{Nd}}-\epsilon_{\text{Sr}}$ diagram showing the hypothetical effects of the experimental metasomatism on upper-mantle harzburgite at 150 Myr BP. All fields are represented for 150 Myr BP except the metasomite field which is shown for the present. Lithosphere: hypothetical garnet peridotite 'protolith'. Metasomatic vectors: **a**, mixing between lithosphere and Group I kimberlite; **b**, time-integrated effect due to Rb/Sr increase; **m**, the resultant metasomatic vector. Data sources: Group I kimberlites¹⁰; garnet peridotites^{1,20}; metasomites⁵. ϵ_{Nd} is the deviation from the value expected in a uniform chondritic reservoir (CHUR) at a time T and is defined as $\epsilon_{\text{Nd}}(T) = 10^4 \{ [(^{143}\text{Nd}/^{144}\text{Nd})_{\text{rock}}(T)] / [(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}] - 1 \}$; ϵ_{Sr} is similarly defined using the isotopes ^{87}Sr and ^{86}Sr .

the combination of the long half-life of ^{147}Sm and the low Sm/Nd produced by the metasomatism will result in only small increases in $^{143}\text{Nd}/^{144}\text{Nd}$ over 150 Myr. Thus the net time-integrated effect of Sm/Nd metasomatism will be a vector of shallow slope towards the 'enriched' quadrant in $\epsilon_{\text{Sr}}-\epsilon_{\text{Nd}}$ space (component **b** in Fig. 2). The combined effect of the two processes is represented by a steep vector towards the upper-right quadrant (component **m**, Fig. 2)—a trend similar to that reported for metasomite xenoliths from southern Africa¹.

The experimentally metasomatized assemblages show a net gain in U, Th and Pb but a decrease in the U/Pb and Th/Pb ratios. Consequently, it may be inferred that increases in $^{208}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ with time will be minor compared to unmetasomatized peridotite and that a negative correlation will develop between strontium and lead isotopic trends. Such a negative correlation has indeed been documented for kimberlite-borne metasomite xenoliths from southern Africa¹. Once again, the trace-element metasomatism shown by the experiment can account for isotopic signatures shown by the natural metasomite xenolith suites.

Many workers consider that the upward passage of low-volume partial melts from the asthenosphere into the overlying lithosphere is an important and widespread mechanism of enrichment of major and trace elements in the lower lithosphere². This enrichment has been variously attributed to a variety of magma types (group I kimberlite⁶, 'kimberlite-related'²² magma, lamproite²³), although there is no clear relationship between the trace-element characteristics of the metasomites and these magmas^{1,4,13}. The experimental results presented here demonstrate that the interaction between a Group I kimberlite melt and refractory mantle can lead to a decoupling of Rb, Pb and K from other trace elements, resulting in selective enrichments of these elements in the metasomites. These results thus support previous arguments, based on isotopic and temporal evidence⁶, that the most recent phase of metasomatism beneath southern Africa is related to Group I kimberlite magmatism. The textural characteristics of the experimental metasomites indicate that this

decoupling is consistent with metasomatism by H_2O -rich fluids generated as a result of crystallization of Group I kimberlite in the lower lithosphere. □

Received 31 May; accepted 14 November 1994.

- Hawkesworth, C. J., Erlank, A. J., Kempton, P. D. & Waters, F. G. *Chem. Geol.* **85**, 19–34 (1990).
- Menzies, M. A. *Nature* **343**, 312–313 (1990).
- McKenzie, D. *Earth planet. Sci. Lett.* **95**, 53–72 (1989).
- Dawson, J. B. & Smith, J. V. *Geochim. cosmochim. Acta* **41**, 309–323 (1977).
- Erlank, A. J. et al. in *Mantle Metasomatism* (eds Hawkesworth, C. J. & Menzies, M. A.) 221–311 (Academic, London, 1987).
- Kinny, P. D. & Dawson, J. B. *Nature* **360**, 726–728 (1992).
- Nixon, P. H. in *Mantle Xenoliths* (ed. Nixon, P. H.) 215–239 (Wiley, Chichester, 1987).
- Harte, B. in *Mantle Xenoliths* (ed. Nixon, P. H.) 625–640 (Wiley, Chichester, 1987).
- Dawson, J. B. *Kimberlites and their Xenoliths* (Springer, 1980).
- Frey, F. & Green, D. H. *Geochim. cosmochim. Acta* **38**, 1023–1059 (1974).
- Smith, C. B. *Nature* **304**, 400–403 (1983).
- Smith, C. B., Gurney, J. J., Skinner, E. M. W., Clemment, C. R. & Ebrahim, N. *Trans. geol. Soc. S. Afr.* **88**, 267–280 (1985).
- Menzies, M. A., Rodgers, N., Tindle, A. & Hawkesworth, C. J. in *Mantle Metasomatism* (eds Hawkesworth, C. J. & Menzies, M. A.) 313–361 (Academic, London, 1987).
- Dawson, J. B. & Hawthorne, J. B. *Bull. volcan.* **34**, 740–757 (1970).
- Cox, K. J., Gurney, J. J. & Harte, B. in *Lesotho Kimberlites* (ed. Nixon, P. H.) 76–91 (Lesotho National Development Corp., Maseru, 1973).
- Eggler, D. H. & Baker, D. R. in *Advances in Earth and Planetary Sciences: High Pressure Research in Geophysics* Vol. 12 (eds Akimoto, S. & Manghani, M. H.) 237–250 (Reidel, Dordrecht, 1982).
- Green, D. H. *Earth planet. Sci. Lett.* **19**, 37–53 (1973).
- Sun, S.-s. & McDonough, W. F. in *Magmatism in Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Spec. Publ. 42, Geol. Soc., Oxford, 1989).
- Bergman, S. C. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Upton, B. G. J.) 103–189 (Spec. Publ. 30, Geol. Soc., Oxford, 1987).
- Kushiro, I. & Erlank, A. J. *Carnegie Inst. Wash. Yb.* **68**, 231–233 (1970).
- Richardson, S. H., Erlank, A. J. & Harte, S. R. *Earth planet. Sci. Lett.* **75**, 116–128 (1973).
- Jones, A. P., Smith, J. V. & Dawson, J. B. *J. Geol.* **90**, 435–452 (1982).
- Bergman, S. C. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Upton, B. G. J.) 103–189 (Spec. Publ. 30, Geol. Soc., Oxford, 1987).
- Waters, F. G. *Contr. Miner. Petrol.* **95**, 523–533 (1987).
- Odling, N. W. A. *Mineralog. Mag.* (in the press).

ACKNOWLEDGEMENTS. I thank J. Dawson, C. Mitchell, S. Harley and J. Gurney for reviews of the manuscript, J. Craven and R. Hinton for assistance with SIMS analysis, and P. Hill and S. Kearns for assistance with electron probe microanalysis. This work was supported by the UK Natural Environment Research Council.

Reproductive constraints on aggressive competition in female baboons

Craig Packer*, D. A. Collins†, A. Sindimwo‡ & J. Goodall‡

* Department EEB, University of Minnesota, 1987 Upper Buford Circle, St Paul, Minnesota 55108, USA

† ICAPB, Ashworth Laboratory, University of Edinburgh, Edinburgh EH9 3JT, UK

‡ Gombe Stream Research Centre, Box 185, Kigoma, Tanzania

COMPETITIVE interaction between females of the same social group is characteristic of most primate species^{1–3}. In Old World monkeys, females of high social rank maintain priority of access to scarce resources and harass low-ranking companions^{1–6}. But different field studies have found differing correlations between female dominance and reproductive success: several populations show an advantage of rank whereas others do not^{1,3,5,7}. Although such variation may reflect divergent levels of predation, food availability or social stress in different environments, female competitive ability may also be balanced by significant reproductive costs and thus be subject to strong stabilizing selection. We report here that high-ranking female baboons (*Papio cynocephalus anubis*) at Gombe National Park, Tanzania, enjoy shorter interbirth intervals, improved infant survival, and accelerated maturation of their daughters. These advantages, however, are countered by a significantly higher probability of miscarriage, and a proportion of high-ranking females suffer from reduced fertility.

Olive baboons at Gombe have been studied continuously since 1967^{8–13}. The National Park consists of a series of steep valleys

that run down the rift escarpment to the shore of Lake Tanganyika¹⁴. The valley floors are filled with semi-evergreen forest, and the slopes and ridgetops are covered by grassland, savanna woodland and semideciduous forest. The dry season at Gombe extends from the beginning of June until the second week of October, and the number of days with rain has been recorded each dry season since 1970. Information is available on 584 pregnancies of 138 females in five troops between 1967 and 1992. Pregnancy can be detected from the colour of the female's perineum¹⁵. The average number of females in each

troop ranged from 11.4 to 17.4 adult females, but none of our results are significantly affected by troop size. Daily demographic data include all dates of birth, death, cycle state, miscarriage, the date of emigration by sons and of menarche in daughters. Annual dominance rankings (based on supplanting from food sites and on dyadic aggressive interactions) are available for about two-thirds of the overall study period; average lifetime ranks are calculated from annual rankings. Out of 141 females ranked over two or more years, 52% varied by fewer than four rank positions over their lifetime, but 12% varied by more than

TABLE 1 Effect of female dominance rank on the major components of fitness

(a) Reproductive performance over the lifetime of individual females					
Dependent variable	Independent variable	<i>t</i>	<i>P</i>	<i>r</i> ²	<i>n</i>
Simple regression					
Average intervals after offspring that survive to two years					
Post-partum amenorrhoea	Average rank	−4.86	0.0000	0.2347	79 females
Cycling→consortship	Average rank	−0.91	0.3632	0.0112	76 females
Consort→conception	Average rank	−1.28	0.2062	0.0215	76 females
Infant survival (for females with two or more live-born offspring)					
Survival in 1st year	Average rank	3.24	0.0017	0.0974	99 females
Survival in 2nd year	Average rank	1.21	0.2288	0.0212	70 females
Lifespan	Average rank	0.67	0.5071	0.0127	37 deceased females
Lifetime reproductive success	Average rank	1.89	0.0675	0.0923	37 deceased females
Daughter's age at first birth	Mother's rank	−1.97	0.0523	0.0493	77 primiparous daughters
Son's age at natal emigration	Mother's rank	−0.09	0.9286	0.0001	88 emigrating sons
Daughter's rank	Mother's rank	9.54	0.0000	0.5323	82 daughters
Son's rank in new troop	Mother's rank	0.45	0.6558	0.0044	48 emigrant sons, 8–14 yr
Multiple regression					
Post-partum amenorrhoea	Average rank	−5.07	0.0000	0.3312	79 females
	A-troop	2.64	0.0102		
	C-troop	2.73	0.0079		
Daughter's age at 1st birth	Mother's rank	−2.27	0.0269	0.5060	62 daughters
	Mother's age	−3.30	0.0017		
	A-troop	5.68	0.0000		
	C-troop	3.65	0.0006		
(b) Intervals after birth of offspring that survived two years					
Simple regression					
Post-partum amenorrhoea	Rank	−5.38	0.0000	0.1424	176 intervals
Cycling→consortship	Rank	−1.08	0.2811	0.0070	167 intervals
Consort→conception	Rank	−0.10	0.9173	0.0001	168 intervals
Multiple regression					
Post-partum amenorrhoea	Rank	−5.75	0.0000	0.4461	160 intervals
	Age	2.19	0.0303		
	A troop	6.58	0.0000		
	C troop	5.31	0.0000		
	Dry-season rain	−3.60	0.0004		
(c) Logistic regression of infant survival (live births)					
Dependent variable	Independent variable	Coefficient/ s.e.m.	<i>P</i>	<i>n</i>	
Simple logistic regression					
Survival in 1st year	Mother's rank	2.70	0.0069	383 infants	
Survival in 2nd year	Mother's rank	0.16	0.8748	245 yearlings	
Multiple logistic regression					
Survival in 1st year	Mother's rank	3.08	0.0021	298 infants	
	First born	−3.17	0.0016		
	Mother's age	−1.91	0.0560		

The significant effects of rank are indicated by bold italics. 'Rank' is the female's rank at the birth of her infant. 'Mother's rank' and 'Average rank' are the female's average rank over her lifetime. 'Reproductive success' is the number of offspring that reached two years of age. Multiple regressions were done whenever additional independent variables significantly influenced the dependent variable. 'Post-partum amenorrhoea' is the interval from birth until the mother shows the first sign of a perineal swelling; 'Cycling→consortship' is the interval from first sign of swelling until the female was sexually consorted by an adult male. Data on the rank of adult sons is restricted to males of 8–14 years of age because of the strongly inverse U-shaped relationship between male rank and age. All data were log-transformed where necessary.

eight positions. By convention, we present correlations with rank as being positive when high-ranking animals have the highest score. Linear and logistic regression are used where appropriate; other tests are non-parametric. All probabilities are two tailed.

High-ranking females at Gombe gain several reproductive advantages: higher infant survival, shorter interbirth intervals following a surviving offspring, and daughters that generally give birth at a younger age (Table 1). We analysed each conception separately as well as the lifetime reproductive performance of individual females because reproductive performance can vary with birth order^{7,16}, age¹⁷ and annual rainfall^{18,19}, and a female's rank can change over the course of her lifetime (see above and ref. 20). All effects of female rank that are significant by simple regression remain significant in the multivariate analyses.

Consistent with many other primate studies^{1-3,5}, the faster reproductive rates of high-ranking Gombe females appear to result from improved nutrition. Post-partum amenorrhoea, which reflects the heavy energetic costs of lactation^{5,16}, is shortest in years with good dry-season rainfall. Post-partum amenorrhoea and daughter's age at first birth both vary significantly across troops, with members of A and C troop taking longest to breed. Males born in A and C troop also emigrate from their natal troop at the oldest ages (A troop: Student's $t = 2.90$, $P = 0.0047$; C troop: $t = 2.80$, $P = 0.0063$; $r^2 = 0.1230$, $n = 88$). Natal emigration by males is restricted to as narrow an age range as the first birth of young females (Fig. 1), providing a similar landmark in physical maturation. Thus low-ranking females in each troop and all the members of A and C troop presumably suffer from an inferior diet. Several studies of captive primates have suggested that low-ranking females suffer reproductive failure due to stress or from harassment from high-ranking rivals^{4,6,21,22}, and longer interbirth intervals by subordinates have also been interpreted as evidence of stress^{4,6,16}. However, once subordinate females resume cycling at Gombe, they become sexually attractive and conceive as rapidly as high-ranking females (Table 1), suggesting that the effects of stress on subordinates may be less severe here than elsewhere.

Although high-ranking Gombe females generally enjoy improved reproductive performance, the overall relationship between lifetime reproductive success and female dominance is not statistically significant (Table 1). One factor that dampens

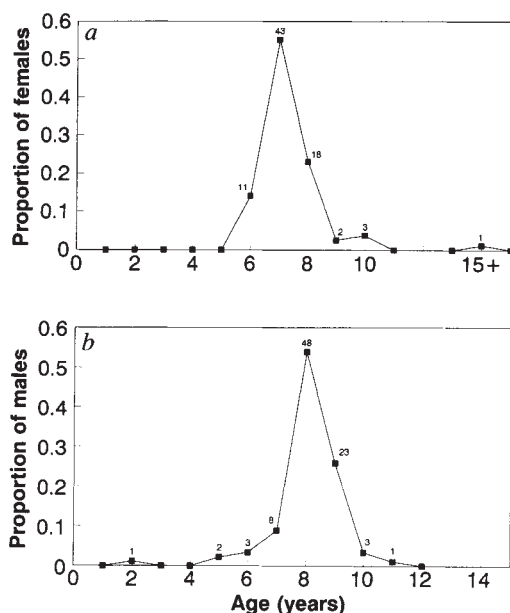


FIG. 1 *a*, Age at first birth for 78 females. Note that one of these females has not yet conceived, even though she is in her 15th year. *b*, Age at natal emigration for 89 males.

TABLE 2 Effect of dominance rank on miscarriage

(a) Miscarriages in individual females

Dependent variable	Independent variable	<i>t</i>	<i>P</i>	<i>r</i> ²	<i>n</i>
Simple regression					
Log number of miscarriages	Average rank	2.47	0.0146	0.0447	133 females
Multiple regression					
Log number of miscarriages	Average rank	1.99	0.0496	0.2666	100 females
	Age at death/end of study	5.42	0.0000		
Spearman correlation					
Proportion of pregnancies miscarried	Average rank	2.06	0.0413	<i>r</i> _s = 0.1779	132 females

(b) Logistic regression of individual pregnancies

Dependent variable	Independent variable	Coefficient/s.e.m.	<i>P</i>	<i>n</i>
Simple logistic regression				
Miscarriage	Rank	3.11	0.0019	467 pregnancies
Multiple logistic regression				
Miscarriage	Rank	2.49	0.0129	288 pregnancies
	Age	3.37	0.0008	
	Survival of prior infant	-2.62	0.0089	

Females of high rank and advanced age have the most miscarriages. Data on number of miscarriages were log-transformed; Spearman correlations were done on the proportion of miscarriages because these data could not be normalized.

the overall trend is the greater incidence of miscarriage in high-ranking females (Table 2). About 10% (56 of 584) of pregnancies end in miscarriage at Gombe, with equal proportions occurring in each trimester. The probability of miscarriage is unaffected by interbirth interval and is higher following the loss of an unweaned infant, thus the effect of dominance is not due to rapid reproductive rate or higher infant survival. Older females are more likely to suffer miscarriage than younger females, but the effect of rank remains significant when (1) age is included in a multivariate analysis (Table 2) or (2) when older females are excluded from the analysis (in a logistic regression, for females ≤ 14 yr old: effect of rank on miscarriage: coefficient/s.e.m. = 2.05, $P = 0.0408$, $n = 263$ pregnancies; female's age and survival of her previous infant no longer significant).

Besides higher rates of miscarriage, a proportion of females from high-ranking lineages also showed signs of reduced fertility. One unusually large and aggressive top-ranking adult female (of unknown parentage) had only one pregnancy, a miscarriage, over her entire life. Four other females showed an unusually late age at first reproduction (Fig. 1a). One of these showed the highest unassisted rank ascent in 62 mother-daughter pairs: whereas most Gombe females attained a rank within three positions of their mother (42 of 62 cases, 67%), this female ascended to second rank whereas her mother ranked eleventh. The remaining three late-maturing daughters were all born to high-ranking females (one mother of second rank, two of third rank), even though daughters of high-ranking females generally gave birth to their first offspring at a younger age (Table 1). Delayed reproduction of these four young females stemmed both from slower maturation and from lowered fertility. All four experienced their first cycle at a greater age than the population average (mean: 6.07 yr, range: 5.3–6.88 yr, $n = 4$; versus population mean: 4.81 yr, range: 3.49–6.25 yr, $n = 90$). All three that eventually gave birth also showed above-average periods of adolescent sterility (mean: 3.35 yr, range: 2.90–4.10 yr; versus population mean: 1.90 yr, range: 1.12–3.28 yr, $n = 74$).

Dominance-related reproductive failure in the Gombe

baboons appears to be 'hit or miss': several high-ranking females raised 8–10 surviving offspring, whereas others raised none. If we exclude all females that never gave birth to live young, the relationship between rank and lifetime reproductive success becomes statistically significant ($t=2.19$, $P=0.0354$, $r^2=0.1272$, $n=35$ deceased females with at least one live birth; multiple regression: effect of rank, $t=2.80$, $P=0.0087$, effect of lifespan: $t=7.60$, $P=0.0000$, overall $r^2=0.6889$, $n=35$). Most primate field studies have focused on only a few social groups for short spans of time. The perceived relationship between rank and reproduction in such studies would be appreciably altered by the presence or absence of even one high-ranking female suffering from reproductive pathology²³.

Our findings suggest that qualities essential to enhancing or maintaining female rank may also carry significant reproductive

costs. High-ranking female baboons have greater difficulties in completing a pregnancy, and a proportion suffer a delay in the onset of reproduction. Similar reproductive pathologies have been observed in other mammalian species where females were exposed to high levels of androgens^{24,26}, and masculinized, infertile females have been observed in several other primate populations (S. K. Wasser, personal communication and ref. 14). Although such females are often viewed as anomalies, reproductive failure may serve as a widespread constraint on traits that promote agonistic competition^{25,26}. Indeed, without such costs, strong directional selection on traits that confer high rank might eventually result in females becoming as aggressive as female spotted hyenas²⁷. The rarity of mammalian species with such hyperaggressive females suggests that these constraints must be formidable²⁶. □

Received 26 September; accepted 28 October 1994.

1. Silk, J. in *Primate Societies* (eds Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T.) 318–329 (Univ. Chicago Press, Chicago, 1987).
2. Walters, J. R. & Seyfarth, R. M. in *Primate Societies* (eds Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T.) 306–317 (Univ. Chicago Press, Chicago, 1987).
3. Chapais, B. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & deWaal, F. B. M.) 29–60 (Oxford Univ. Press, Oxford, 1992).
4. Wasser, S. K. in *Social Behavior of Female Vertebrates* (ed. Wasser, S. K.) 349–390 (Academic, New York, 1983).
5. Lee, P. C. *J. Zool.* **213**, 409–422 (1987).
6. Abbott, D. H. *J. Zool.* **213**, 455–470 (1987).
7. Altmann, J., Hausfater, G. & Altmann, S. A. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 403–418 (Univ. Chicago Press, Chicago, 1988).
8. Packer, C. *Nature* **255**, 219–220 (1975).
9. Nash, L. T. *Am. J. phys. Anthropol.* **44**, 63–77 (1976).
10. Owens, N. W. *Behaviour* **57**, 241–259 (1976).
11. Packer, C. *Anim. Behav.* **27**, 1–36 (1979).
12. Ransom, T. W. *Beach Troop of the Gombe* (Associated University Presses, New Brunswick, NJ, 1981).

13. Collins, D. A., Busse, C. D. & Goodall, J. in *Infanticide* (eds Hausfater, G. & Hardy, S. B.) 193–216 (Aldine, Hawthorne, NY, 1984).
14. Goodall, J. *The Chimpanzees of Gombe* (Harvard Univ. Press, Cambridge, MA, 1986).
15. Altmann, S. A. *Lab. Anim. Digest* **6**, 7–10 (1970).
16. Smuts, B. B. & Nicolson, N. D. *Am. J. Primatol.* **19**, 229–246 (1989).
17. Strum, S. C. & Western, J. *Am. J. Primatol.* **3**, 61–76 (1982).
18. Bercovitch, F. B. & Strum, S. C. *Behav. Ecol. Sociobiol.* **33**, 313–319 (1993).
19. Wasser, S. K. & Norton, G. *Behav. Ecol. Sociobiol.* **32**, 273–281 (1993).
20. Hausfater, G., Altmann, J. & Altmann, S. A. *Science* **217**, 752–755 (1982).
21. Sackett, G. P., Holm, R. A., Davis, A. E. & Fahnenbuck, E. E. in *Proc. 5th Congr. Int. Primatological Society* (eds Kondo, S., Kawai, M., Ehara, A. & Kawamura, S.) (Japan Science, Tokyo, 1975).
22. Fairbanks, L. A. & McGuire, M. T. *Anim. Behav.* **34**, 1710–1721 (1986).
23. Hausfater, G. *Dominance and Reproduction in Baboons* (Karger, Basel, 1975).
24. Goy, R. W. & Resko, J. A. *Recent Progr. Horm. Res.* **28**, 707 (1972).
25. vom Saal, F. J. *Reprod. Fert.* **86**, 457–471 (1989).
26. Glickman, S. E., Frank, L. G., Holekamp, K. E., Smale, L. & Licht, P. *Perspect. Ethol.* **10**, 87–117 (1993).
27. Yalcinkaya, T. M. *et al. Science* **260**, 1929–1931 (1993).

ACKNOWLEDGEMENTS. We thank the Government of Tanzania for permission and A. E. Pusey, F. vom Saal and S. K. Wasser for comments and discussion. Long-term studies at Gombe are supported by the Jane Goodall Institute and the NSF.

Implications for lateralization of bird song from unilateral gating of bilateral motor patterns

Franz Goller* & Roderick A. Suthers*†

* School of Medicine and † Program in Neural Sciences, Myers Hall, Indiana University, Bloomington, Indiana 47405, USA

FUNCTIONAL lateralization of the brain, once considered unique to human language, remains poorly understood^{1–5}. The most convincing example of this phenomenon in animals is singing behaviour in songbirds^{1,4}, in which asymmetries range from a clear unilateral dominance^{6–8} to approximately equal contributions from each side of the vocal organ, the syrinx⁹. Here we report that in brown thrashers (*Toxostoma rufum*) only the activity of muscles that gate sound production by regulating airflow through each side of the syrinx is lateralized. Other syringeal muscles that primarily control the phonetic structure of vocalizations are active on both sides of the syrinx. This explains the puzzling absence of laterality in the morphology^{10,11} and activity^{12,13} of the higher central song control nuclei and suggests that song lateralization did not evolve as a means of achieving a single 'executive' command centre, or as a way of economizing on motor circuits to free brain space for other tasks^{14–16}.

In brown thrashers, both sides of the syrinx contribute approximately equally to song production. Although neither side of the thrasher syrinx dominates song production, the motor control of song is lateralized in the sense that each side can generate sound independently from the other, regardless of

whether phonation occurs sequentially on each side or simultaneously on both sides of the syrinx⁹. We recorded the electromyograms (EMGs) of six syringeal muscles (up to four muscles at one time in each of nine adult males) (Fig. 1) together with the rate of airflow through each side of the syrinx, the subsyringeal pressure, and the vocal output in singing brown thrashers.

The gating of sound production on each side of the syrinx is primarily controlled by the action of the dorsal muscles, m. syringealis dorsalis and m. tracheobronchialis dorsalis, which increase the ipsilateral syringeal resistance, presumably by adducting the lateral labium into the syringeal lumen. Complete closure of either side of the syrinx silences it by preventing expiratory airflow. Through their unilateral adductive action, as indicated by increased resistance to syringeal airflow, these muscles thus determine the extent to which each side of the syrinx contributes to the song. The silencing, by closure, of one side of the syrinx is always preceded (9.2 ± 0.3 ms) and accompanied by electrical activity in the ipsilateral m. syringealis dorsalis and m. tracheobronchialis dorsalis (Fig. 2a), which are also the only syringeal muscles whose electrical activity consistently coincides with increases in ipsilateral syringeal resistance (Fig. 2b). Lateralized song production depends on the unilateral activation of these muscles which, by controlling airflow through the ipsilateral syrinx, permit the independent gating of song on each side of the vocal organ. During phonation these dorsal muscles may also affect the properties of the sound, but their role in gating sound production seems to be unique.

The activity of the other syringeal muscles is not correlated with adduction and is independent of whether or not the ipsilateral side of the syrinx is producing sound (Figs 2, 3). These muscles appear to be involved in controlling the phonology of the ongoing syllable. This is certainly true for the largest of them, m. syringealis ventralis, which has a major role in regulating

FIG. 1 Ventro-lateral view of the syrinx of a brown thrasher, showing syringeal muscles. TL, musculus tracheolateralis; ST, m. sternotrachealis; vTB, m. tracheobronchialis ventralis; dTB, m. tracheobronchialis dorsalis; vS, m. syringealis ventralis; dS, m. syringealis dorsalis; T, trachea; B, bronchial cartilage; ICM, membrane of interclavicular air sac. The oscine syrinx is a bipartite structure at the base of the trachea and cranial end of the primary bronchi. Each bronchus contains putative sound-generating medial tympaniform membranes and a lateral labium which can be adducted into the lumen to act as a valve controlling the rate of airflow across the medial tympaniform membrane. In all song-birds studied, the motor output to the syrinx arises in the efferent pathway of a well developed system of song control nuclei¹⁸. Bilateral connections between these central nuclei have not been demonstrated above the thalamus. Motor neurons in the tracheosyringeal portion of each hypoglossal nucleus innervate the ipsilateral members of syringeal muscles.

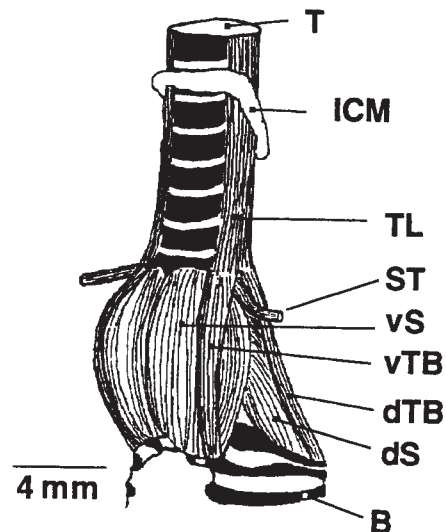
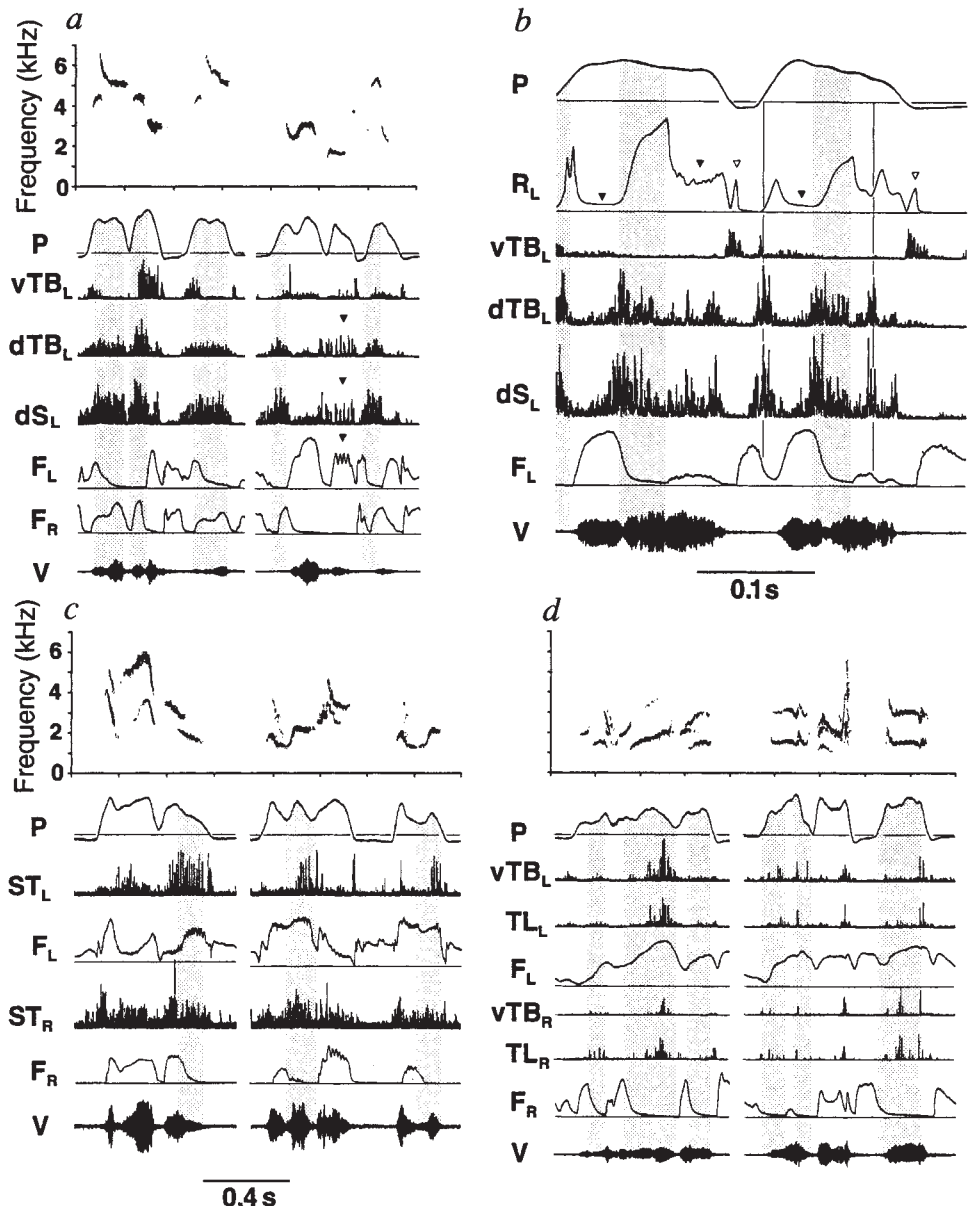


FIG. 2 Only the dorsal muscles, dTB_L and dS_L, have high-amplitude EMGs closely associated with adduction. *a*, Activity of left dorsal muscles (dTB_L and dS_L) is high during adduction (shaded columns) of the left syrinx. In a syllable with rapid modulations of flow rate, distinct bursts of electrical activity in the dorsal muscles precede each decrease in flow (arrowheads). *b*, Electrical activity in dorsal muscles correlates with syringeal resistance (R_L ; horizontal line indicates zero resistance, shading as in *a*). EMG spikes in dorsal muscles generally precede increasing resistance (transient changes indicated by vertical lines). EMG response is proportional to sustained high resistance (filled triangles). During the switch from expiratory to inspiratory airflow (air sac pressure is negative), resistance briefly increases (open triangles). *c*, *d*, EMG patterns of other syringeal muscles (ST, vTB, TL) are not consistently correlated with periods of ipsilateral adduction (shaded columns mark right-side adduction) and must have another role in phonation. Vocalizations are represented oscillographically (V) and spectrographically (top panel in *a*, *c*, *d*); electromyograms are rectified. Symbols for muscles are as for Fig. 1; F_L , rate of airflow; P , air sac pressure (horizontal lines marks zero pressure and flow); subscripts depict distal side of the syrinx (L, left; R, right); R_L , left syringeal resistance is approximated by $R_L = P/F_L$. Timescale is the same in *a*, *c* and *d*. Methods for recording subsyringeal pressure and bilateral airflow have been described^{9,17}. Electromyograms were recorded with bipolar stainless steel wire electrodes ($\phi = 25 \mu\text{m}$, insulated except for 0.2 mm at the tip). Signals were differentially amplified (Princeton Applied Research model 113, band-pass filters set to 0.2–3 kHz) and digitally recorded together with the other parameters. Interference from cross-talk was found to be insignificant based on visual inspection and cross-correlation of EMGs¹⁹.



sound frequency, presumably by controlling the tension on the medial tympaniform membranes. The amplitude of its electrical activity has a strong positive correlation with the fundamental frequency but there is only a weak correlation between frequency and resistance (Fig. 4a, b). This correlation is particularly striking in syllables having pronounced frequency modulation (Fig. 4c).

Both *m. syringealis ventralis* receive motor excitation during phonation, regardless of which side of the syrinx is generating sound, and the ventralis activity on the silent adducted side parallels that on the open phonating side. The strength of the EMG

is also proportional to the fundamental frequency of the sound (Fig. 4b).

Furthermore, the correlation between ventralis electrical activity and contralateral sound frequency reflects the different average fundamental frequency ranges of the two sides of the

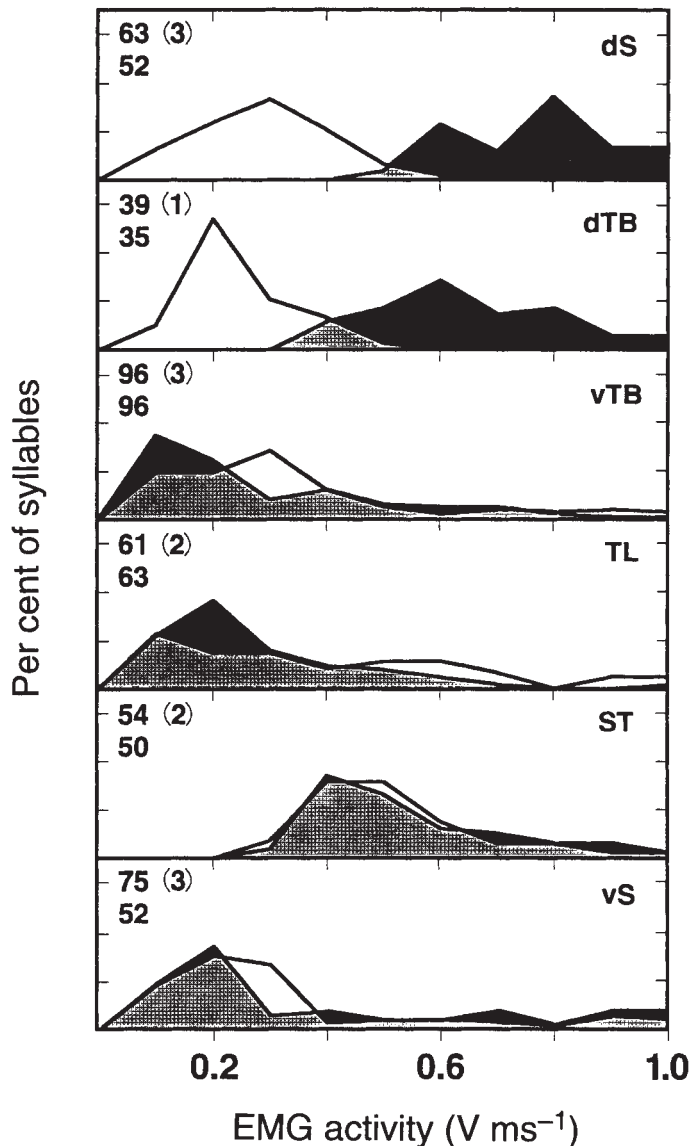


FIG. 3 Frequency distributions of the normalized amplitude of EMG activity in syringeal muscles (symbols as for Fig. 1) during full adduction while the contralateral side generates sound (black area) and during ipsilaterally produced syllables (white area). The overlapping area of the two distributions is shaded. The difference in distributions is highly significant for dS and dTB ($P < 0.001$), marginal for TL ($P < 0.05$) and not significant for all other muscles ($P > 0.05$) (homogeneity test of independent samples according to Kolmogorov and Smirnov²⁰). In each panel, the number of syllables measured during adduction (upper number) and phonation (lower number) is indicated. The number of individuals contributing to the data sets for each muscle is given in parentheses.

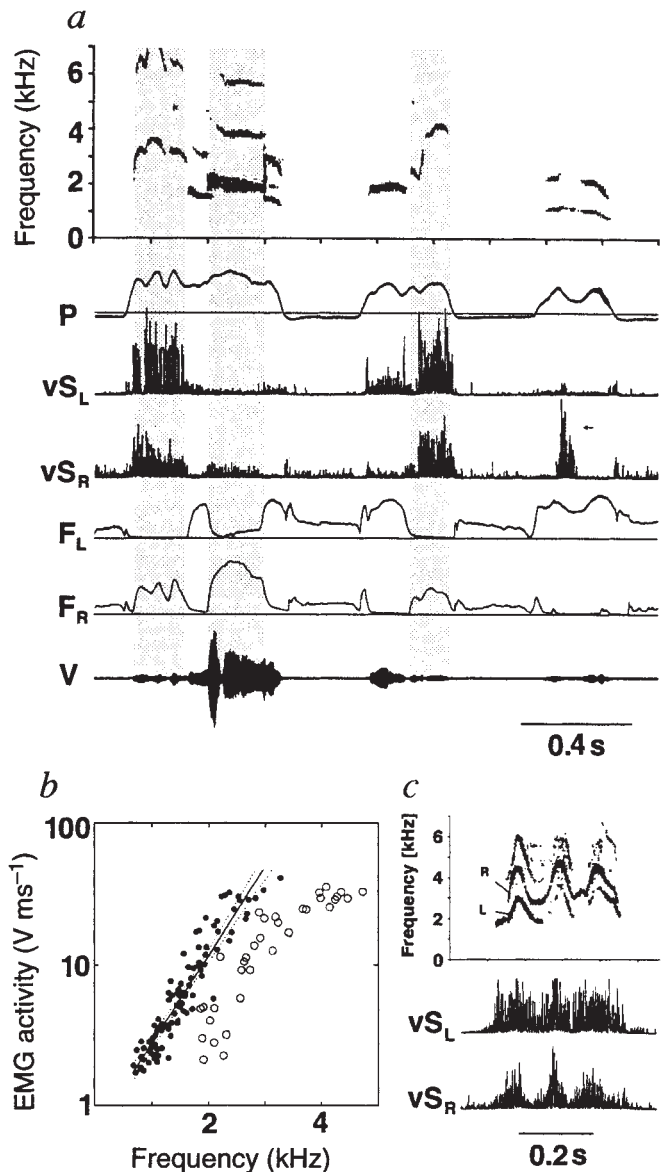


FIG. 4 EMG activity in *m. syringealis ventralis* is closely correlated with the fundamental frequency of the vocalization. a, Bilateral activity in *m. syringealis ventralis* during syllables produced alternately on each side of the syrinx. Shaded areas mark syllables generated on the right side; that is, the left side is closed. Other symbols as for Figs 1 and 2. High amplitude EMGs are associated with higher-frequency notes. The arrow marks an EMG burst in vS_R that is not accompanied by an increase in frequency of the vocal output as the right side is closed. b, Relationship between EMG activity in ventralis muscle and the mean fundamental frequency of the vocalization. This representative example shows data from the left ventralis muscle of one individual. EMG strength increases exponentially with rising fundamental frequency of ipsilaterally (filled circles; regression line $\pm 99\%$ confidence intervals, $R^2 = 0.92$) and contralaterally (the ipsilateral side is closed) (open circles) generated sounds. Correlations of sound frequency and syringeal resistance (data not shown) are considerably weaker (R^2 ranges in 3 birds from 0.06 to 0.25), indicating that the relationship between frequency and EMG strength is not caused by covariance with resistance. c, Example of the close relationship of EMG patterns in ventralis muscles and frequency modulation of the sound.

syrinx, in which the right side tends to generate frequencies 0.5–1.0 kHz higher than the left¹⁷. Thus, the regression line relating fundamental frequency to EMG activity is shifted to the right, so that at any amplitude of EMG the fundamental frequency of the right syrinx is 0.5–1.0 kHz higher than that in the left syrinx.

This selective expression of bilateral motor patterns by unilateral gating in brown thrashers presents a novel mechanism for neural lateralization. It can also explain the paradox of highly asymmetrical song production in the absence of hemispheric differences in the size, neuronal structure^{10,11} or activity^{12,13} of the song control nuclei in other songbirds¹⁶. Bilateral activity of the non-gating muscles may stabilize the cartilaginous framework and pretune the silent side to allow rapid switching between the two sides, accounting for bilateral central motor activity during asymmetrical sound generation^{12,13,16}. Thus, lateralization of song production may have evolved to increase the spectral and temporal complexity of signals used in vocal communication. Conventional hypotheses explaining song lateralization as a means of achieving a unified control centre or to free brain space for other tasks, such as additional memory^{14,16}, need to be re-evaluated. □

Received 15 April; accepted 4 November 1994.

1. Glick, S. D. (ed.) *Cerebral Lateralization in Nonhuman Species* (Academic, New York, 1985).
2. Hellige, J. B. *Hemispheric Asymmetries: What's Right and What's Left* (Harvard Univ. Press, Cambridge, 1993).
3. Ward, D. P. & Hopkins, W. D. *Primate Lateralization* (Springer, New York, 1993).
4. Kimura, D. *Neuromotor Mechanisms in Human Communication* (Oxford Univ. Press, New York, 1993).
5. Corballis, M. C. *The Lopsided Ape* (Oxford Univ. Press, New York, 1991).
6. Nottebohm, F. *Science* **167**, 950–956 (1970).
7. Nottebohm, F. *J. exp. Zool.* **177**, 229–262 (1971).
8. Nottebohm, F. in *Lateralization in the Nervous System* (eds Harnard, S., Doty, R. W., Goldstein, L., Jaynes, J. & Krauthamer, G.) 23–44 (Academic, New York, 1977).
9. Suthers, R. A. *Nature* **347**, 473–477 (1990).
10. De Voogd, T. J. & Nottebohm, F. *J. comp. Neurol.* **196**, 309–316 (1981).
11. Nottebohm, F., Kasparian, S. & Pandazis, C. *Brain Res.* **213**, 99–109 (1981).
12. McCasland, J. S. thesis, California Inst. Technology (1983).
13. McCasland, J. S. *J. Neurosci.* **7**, 23–29 (1987).
14. Nottebohm, F. in *Cerebral Dominance* (eds Geschwind, N. & Galaburda, A.) 93–113 (Harvard Univ. Press, Cambridge, 1984).
15. Nottebohm, F. in *Handbook of Behavioral Neurobiology 2: Neuropsychology* (ed. Gazzaniga, M. S.) 295–344 (Plenum, New York, 1979).
16. Zaidel, E. in *The Dual Brain: Hemispheric Specialization in Humans* (eds Benton, D. F. & Zaidel, E.) 47–63 (UCLA Forum in Med. Sci. No. 26, The Guilford Press, New York, 1985).
17. Suthers, R. A., Goller, F. & Hartley, R. S. *J. Neurobiol.* **25**, 917–936 (1994).
18. Nottebohm, F., Stokes, T. M. & Leonard, C. M. *J. comp. Neurol.* **165**, 456–486 (1976).
19. Loeb, G. E. & Gans, C. *Electromyography for Experimentalists* (Univ. Chicago Press, 1986).
20. Sachs, L. *Angewandte Statistik* (Springer, Berlin, 1984).

ACKNOWLEDGEMENTS. This work was supported by the NIH and the Center for the Integrative Study of Animal Behavior at Indiana University. We thank S. Allan, C. Borland, H. R. Lezotte and S. Rich for assistance and A. Feng, M. Hauser, M. Konishi, F. Nottebohm, J. Phillips and A. Popper for critical comments.

Visual attention modulates metacontrast masking

V. S. Ramachandran & S. Cobb

Brain and Perception Laboratory, Center for Research on Brain and Cognition, 0109, University of California at San Diego, La Jolla, California 92093-0109, USA

How does the human visual system 'bind' different fragments in the visual scene to create enduring representations of objects^{1–5}? A visual illusion known as 'metacontrast'^{6–9} or backward masking provides compelling evidence that perception is not instantaneous and that it occurs sequentially in distinct stages. If a solid white target square is displayed for 50 ms in a tachistoscope, switched off, and followed by a 50 ms display of two flanking mask squares, remarkably, subjects report seeing only the two flanking squares: the first square is simply not 'seen'. By plotting the magnitude of masking as a function of the delay between the target and mask (the stimulus onset asynchrony), one can obtain a characteristic 'U'-shaped function⁷ with optimum masking occurring at about 50 ms, and no masking with synchronous target and mask presentations or at delays higher than 300 ms. The illusion is also highly sensitive to elementary stimulus dimensions such as colour, orientation and spatial frequency⁸, and it has been suggested¹⁰ that it is based on 'low level' autonomous visual mechanisms rather than cognitive processes. Here we describe a novel visual stimulus that demonstrates that metacontrast can be strongly modulated by 'top down' influences such as voluntary visual attention.

Our stimulus sequence was identical to a standard metacontrast display (Fig. 1a) except that our central target was a disk rather than a square. The target disk (a) in Fig. 1b was presented for 100 ms in frame 1 and followed by two flanking squares (s) that appeared in frame 2 for another 100 ms. This caused a complete erasure of the target from consciousness. We then added two additional disks in frame 1 (Fig. 1b), one adjacent to the target (b) and a second one (c) off to one corner of the screen. The stimulus duration and the stimulus onset asynchrony (SOA) (both 100 ms) were longer than in the standard metacontrast display. This stimulus sequence was cycled continuously with a long inter-cycle interval of 700 ms. While viewing this display, we noticed that masking of the target (a in Fig. 1b) occurred as expected when we perceptually 'grouped' the two

extraneous disks (b) and (c). But we found that if we now grouped target (a) and disk (b) instead, the masking no longer occurred: the target became clearly visible. To confirm this observation, we asked four subjects to view this display and to group perceptually either target (a) and disk (b) or disks (b) and (c) (on different trials) and to rate the vividness and clarity of the target on a scale of 0–5 in relation to the flanking squares while we varied the SOA. A characteristic U-shaped metacontrast masking function (Fig. 2) was obtained, but masking was reduced considerably when the subjects grouped the two disks (a) and (b) instead of (b) and (c). This difference between the two conditions disappears if the delay between target and mask is either too long or too short (the two sides of the 'U') suggesting that the effect arises from an interaction between backward masking and visual attention rather than from either alone. We may conclude, therefore, that when visual attention is used to bind the target disc with other adjacent features in the image, masking is reduced considerably.

In our second experiment the stimulus sequence was again similar to a standard metacontrast display, except that the two new discs appeared at the same time as the central target, one on either side of it, and two new squares appeared at the same time as the flanking mask squares (Fig. 3). The stimuli were cycled continuously with an interval of 700 ms interposed between successive target-mask sequences. In this display, when we attended to the vertical column of squares, the middle disk (the target) was masked vividly, as expected, but when we attended to the horizontal row, the masking was reduced considerably, so that three disks were seen instead of two. By voluntarily switching attention from vertical column to horizontal row, we could make the target disk disappear and reappear at will.

An alternative view of metacontrast is that it arises in close association with the apparent motion⁷ between the target and mask, that is, the mechanism of apparent motion that extracts the motion signal after both stimuli have been displayed does not allow the first stimulus to be recognized as a separate event (as the same object cannot be at two places at the same time). Because voluntary attention is known to modulate long-range apparent motion^{11–14}, it is conceivable that it reduces backward masking as well. In our experiment, the attention-based binding of the target with other disks in frame 1 might reduce the strength of the motion signal from the target to the flanking masks, thereby also reducing the strength of metacontrast mask-

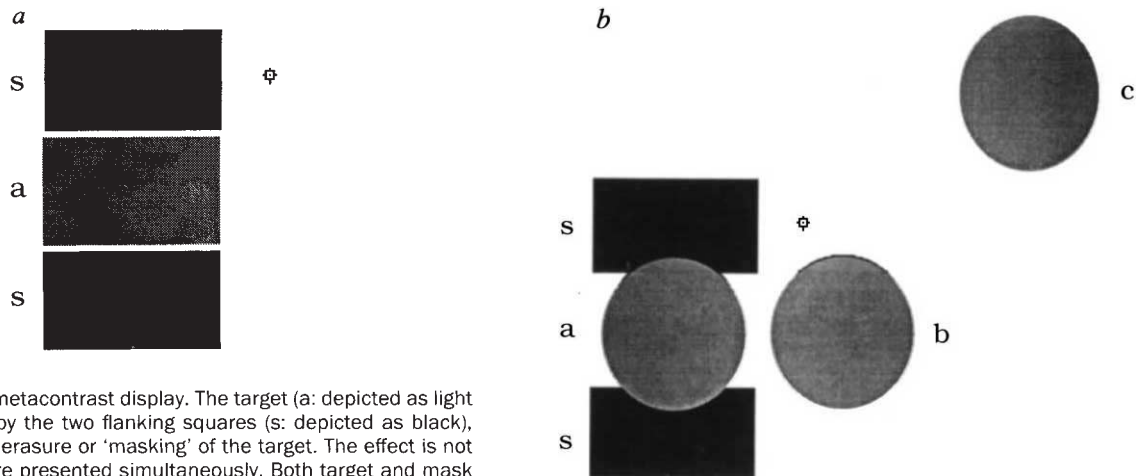


FIG. 1 *a*, Standard metacontrast display. The target (*a*: depicted as light grey) was followed by the two flanking squares (*s*: depicted as black), causing a complete erasure or 'masking' of the target. The effect is not seen if (*a*) and (*s*) are presented simultaneously. Both target and mask were white in the actual display. *b*, Modified display used to demonstrate the effect of attention and perceptual grouping on metacontrast. The display is similar to *a*, except that the target is a disk rather than a square. Also, two new disks, (*b*) and (*c*), were displayed simultaneously with the target. The masking squares (*s*) subtended 0.5° each and the diameter of the disks was also 0.5° . The outer margin of the target disk (*a*) coincided spatially with the lower margin of the upper square and

upper margin of the lower square. Disks (*a*) and (*b*) were 0.8° apart (centre to centre) and disk (*c*) was 2° away from the target, towards the corner of the display. The durations of both target and mask were 100 ms, SOA was 125 ms. To minimize eye movements, the subject was asked always to fixate on a small cross.

FIG. 2 Modulation of masking by visual attention. Durations of both target and mask were 100 ms. On any given trial, SOA was randomly chosen to be one of 6 different values. On different trials, the subject had to 'group' either (*a*) and (*b*), or (*b*) and (*c*) of Fig. 1*b*, and rate the perceptual clarity of the target (*a*) on a scale of 0 to 5 (0, invisible; 5, perfectly clear). Trials on which (*a*) and (*b*) were grouped were interleaved with trials in which (*b*) and (*c*) were grouped. On any given trial and target-mask sequence was followed by a long delay of 600 ms before repeating the same sequence and the stimulus was cycled continuously. This gave the subject enough time to group the disks. No time limit was specified but the subject was encouraged to respond quickly (typically 4 or 5 cycles were required to make a confident judgement). Each data point is based on 24 trials (6 subjects, 4 trials each). The square symbols (■) depict the masking when subjects grouped (*b*) and (*c*) and the circular symbols (●) when they grouped (*a*) and (*b*). Note the striking reduction in masking when (*a*) and (*b*) were grouped at intermediate SOAs and the absence of a difference in masking strength when the SOA was either zero or larger than 160 ms.

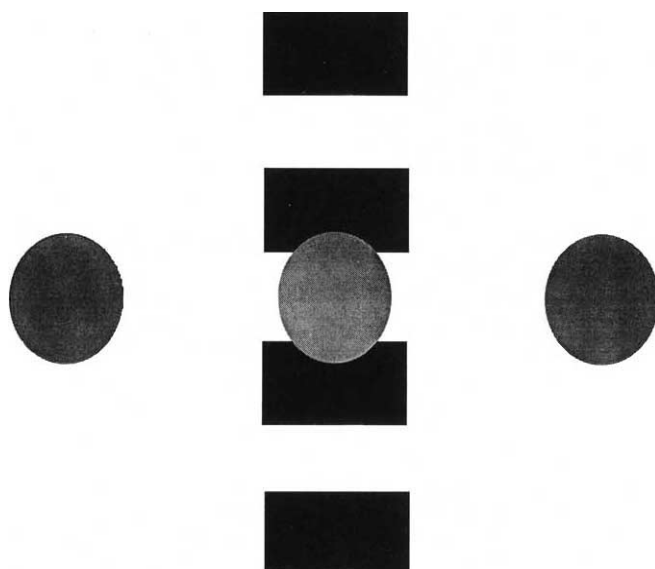
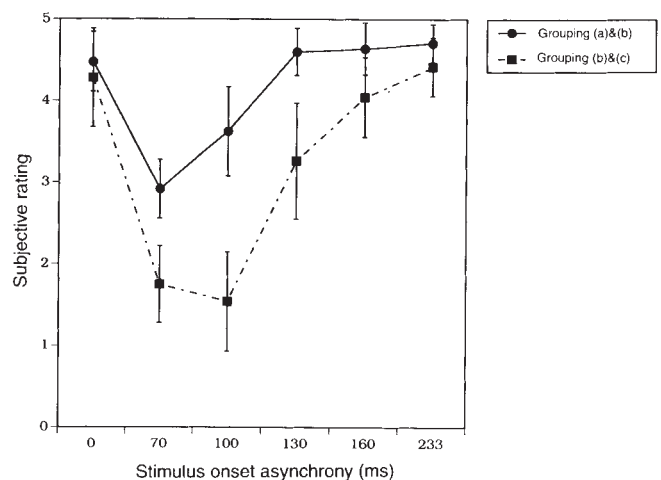


FIG. 3 The display used in experiment 2. In frame 1 we displayed a horizontal row of three white disks and in frame 2 there was a vertical column of four white squares. Paying attention to the mask (the vertical column) caused the central target to disappear, but attending to the horizontal row caused the central disk to become unmasked so that three disks were seen instead of two. We used this display to test five subjects (including the two authors). The squares subtended 0.5° each and the stimulus duration and SOA were 100 ms and 116 ms, respectively. The long inter-cycle interval between successive target-mask pairs was 700 ms. In any given trial, the subject was initially instructed to pay attention to the vertical column of figures, and to rate the clarity of the central disk on a scale of 0–5 (0, visible; 5, clearly visible). He was not given a time limit but was encouraged to respond quickly. The process was repeated while the subject attended to the horizontal row. We also randomly varied the colour of the display so that on different trials all figures were a light desaturated pink, green, or white. This procedure had no noticeable effect on the illusion but it conveyed the impression to the three naive subjects that we were varying something. On 50% of trials the target followed the mask instead of preceding it. The mean ratings for horizontal and vertical were 4.08 and 1.17, respectively ($n = 180$; 5 subjects $\times$ 36 trials). The ratings for the forward masking condition were 4.6 (horizontal) and 4.02 (vertical). This confirmed our preliminary observation that masking is reduced considerably when the subject attends to the horizontal row even though the physical stimulus remains unchanged. The effect of attention could be seen for both backward and forward masking but is much more pronounced for the former.

ing. Indeed, our subjects reported that in the display illustrated in Fig. 1b the target (a) could sometimes be seen 'splitting' into the two masking squares, despite the shape difference (even though the target itself was rendered invisible) but that this splitting did not occur if target (a) was grouped perceptually with disk (b).

When discrete stimuli excite the retina in rapid temporal and spatial proximity, decisions have to be made by a neural network about both the direction of motion and the object identity, with the latter process leading to metacontrast. One would expect, therefore, that although apparent motion and metacontrast are not identical, they should nonetheless be closely linked, and may be sensitive to the same temporal and spatial parameters. To explore this link more directly, we designed a new display (Fig. 4a) in which frame 1 contained two extra squares that appeared adjacent (in space) to the two masking squares that followed in

frame 2. Because of their proximity and shape similarity, subjects now saw vivid apparent horizontal motion between the squares in frame 1 and those in frame 2. This reduced the tendency to see the target disk splitting and also reduced backward masking considerably (see Fig. 4b legend).

These results show that far from being an autonomous front-end visual process, backward masking can be strongly modulated by top-down influences such as voluntary visual attention. The effect we have observed may therefore provide a new way of linking attention, motion perception and immediate memory, and the technique may also lend itself to physiological experiments on visual attention¹⁵⁻¹⁷.

Note added in proof: Unmasking by voluntary attention, at long SOAs, was especially salient with Fig. 1b, but could occasionally be observed if the two extraneous disks (b and c) were deleted so as to exclude perceptual grouping. □

Received 1 September; accepted 13 October 1994.

1. Treisman, A. *Sci. Amer.* **225**, 114-126 (1986).
2. Crick, F. H. C. & Koch, C. *Sem. Neurosci.* **2**, 263-275 (1990).
3. Gray, C. M. & Singer, W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1698-1702 (1989).
4. Damasio, A., Damasio, H., Tranel, D. & Brandt, S. *Cold Spring Harbor Symp.* **55**, 1039-1067 (1990).
5. Finkel, L. & Saida, P. *Amer. Sci.* **82**, 224-237 (1994).
6. Stigler, R. *Pflügers Archiv.* **134**, 365-435 (1910).
7. Kahneman, D. *Psychol. Bulletin* **62**, 404-425 (1968).
8. Williams, M. C., Breitmeyer, B., Lovegrove, W. & Gutierrez, L. *Vis. Res.* **31**, 2017-2023 (1991).
9. Dennett, D. *Consciousness Explained* (Little, Brown, Boston, 1991).
10. Breitmeyer, B. & Ganz, L. *Vis. Res.* **17**, 861-865 (1977).
11. Cobb, S., Ramachandran, V. S. & Rogers-Ramachandran, D. *Soc. Neurosci. Abstr.* **20**, 1667 (1994).
12. Ramachandran, V. S. & Anstis, S. M. *Nature* **304**, 529-531 (1983).
13. Ramachandran, V. S. & Anstis, S. M. *Scient. Am.* **254**, 102-109 (1986).
14. Cavanagh, P. *Science* **257**, 1563-1565 (1992).
15. Moran, J. & Desimone, R. *Science* **229**, 782-784 (1985).
16. Logothetis, N. & Schall, J. *Science* **245**, 761-763 (1989).
17. DiLollo, V., Bischof, W. F. & Dixon, P. *Psychol. Sci.* **4**, 260-263 (1993).

ACKNOWLEDGEMENTS. We thank F. H. C. Crick, H. Pashler, P. Churchland and J. Ramachandran for stimulating discussions.

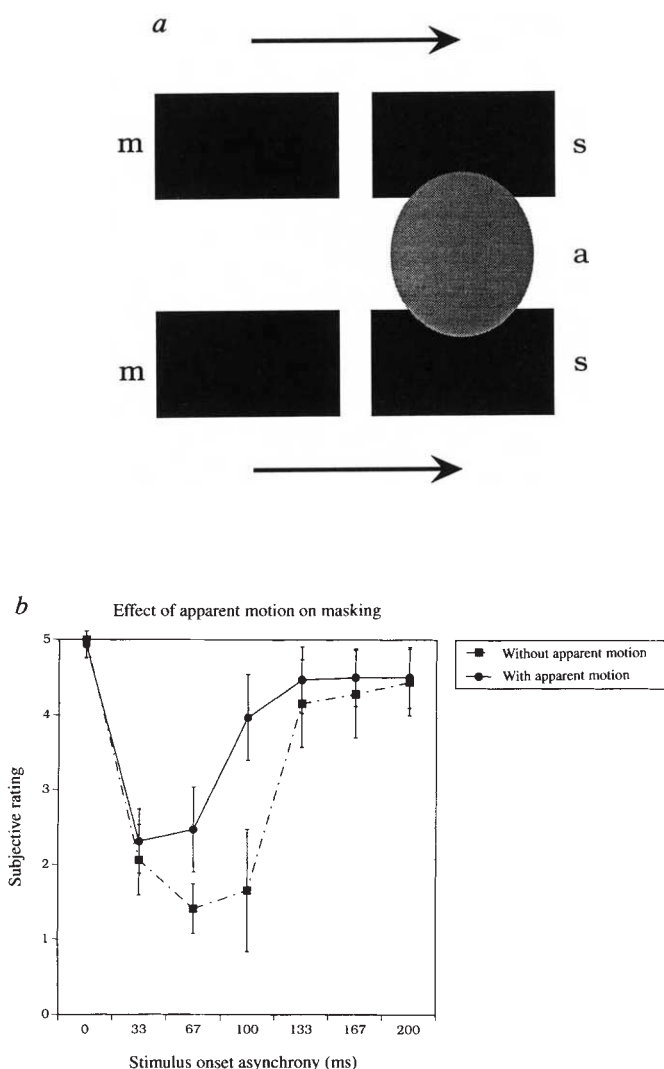


FIG. 4 a, If the flanking squares (s) are made to participate in apparent motion with a different set of features (m) in frame 1, then the masking of the central target (a) in that frame is reduced considerably. The proximity of the two new squares (m) that have been added in frame 1 to the mask squares in frame 2 causes preferential perception of horizontal apparent motion (arrows) so that the target is no longer seen to split. This causes an unmasking of the target (a). b, A formal experiment was conducted on 8 subjects. In 50% of the trials the two extra squares were added to frame 1, whereas in the remaining trials only the target disk appeared in that frame. The SOA was also randomized on different trials, as in experiment 1. Subjects were asked to rate the vividness of the central target disk on a scale of 0 to 5. A significant 'unmasking' of the target occurs (circles) when the extra squares are added ($n = 32$ trials, 8 subjects $\times$ 4 trials).

Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish

Manfred Scharl^{*}, Indrajit Nanda[†], Ingo Schlupp[‡], Brigitta Wilde^{*}, Jörg T. Epplen[§], Michael Schmid[†] & Jakob Parzefall[‡]

^{*} Physiologische Chemie I and [†] Institut für Humangenetik, Biozentrum der Universität, Am Hubland, D-97074 Würzburg, Germany

[‡] Zoologisches Institut und Zoologisches Museum der Universität, Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany

[§] Molekulare Humangenetik der Ruhr-Universität, D-44780 Bochum, Germany

A CENTRAL paradigm in evolutionary biology is that sexual reproduction is advantageous over asexuality¹⁻⁵. One of the long-term disadvantages asexual forms have to face is Muller's ratchet⁶. In the absence of recombination, theoretically no genotype can ever produce offspring with fewer mutations than its own load. The accumulation of deleterious mutations and gene combinations that cannot be purged should lead to extinction of parthenogenetic forms within 10^4 - 10^5 generations^{7,8}. Evidence is accumulating, however, that some of these might have survived for such periods or even longer⁹⁻¹⁴. In the Amazon Molly fish *Poecilia formosa* we have detected a process that appears to compensate for disadvantages of asexuality, namely incorporation of subgenomic amounts of DNA from a bisexual host species by microchromosomes.

The Amazon Molly fish, *P. formosa*, is an all-female gynoform¹⁵ that depends on sperm of males of related bisexual species, *P. mexicana* or *P. latipinna*, for gynogenesis, that is, the sperm is needed to trigger embryogenesis physiologically (reviewed in ref. 16). Eggs are formed without meiosis¹⁷, leading to clonal lineages of genetically identical fish. Thus all the evidence so far supports the concept that males do not contribute genetic material.

Amazon Mollies (Fig. 1a) are usually bred in the laboratory with males of an ornamental, homogeneously black-coloured *Poecilia* (so-called Black Mollies; Fig. 1b). Their pigmentation is due to a specific cell type, the macromelanophore, which is easily distinguished by its enormous size from micromelanophores making up the uniform light-grey wild-type coloration. The macromelanophore pigmentation is encoded by dominant genes, present in the Black Molly, but absent from *P. formosa*¹⁸. In such laboratory broods *P. formosa* generally have wild-type pigmentation owing to gynogenesis. In rare cases (frequency $\sim 10^{-3}$), however, offspring appear that deviate from their homogeneously greyish to olivaceous pigmented sisters. They have patches of black pigmentation of Black Molly type macromelanophores (Fig. 1c). The pattern is clearly different from the

evenly distributed spotting of F₁ hybrids from crosses of Black Mollies with any light-pigmented bisexual *Poecilia* (see Fig. 1d). All such unusually pigmented *P. formosa* reported so far did not transmit the pigmentation phenotype to the offspring¹⁹. They were like normal *P. formosa* cytogenetically and with respect to genome size^{20,21}. We obtained two black-patched animals out of 14, which transmitted the black pigmentation to the following generations, one of which was genetically analysed in more detail. The offspring phenotypes ranged from light to heavily spotted (Fig. 1e,f); even totally black animals were obtained (Fig. 1g). When mated either to Black Molly males or to the light-pigmented males of the natural host species *P. mexicana*, both spotted and black females produced wild-type-pigmented, spotted and homogeneously black offspring (Table 1). In total, the black phenotype was rare (2–3%). Light-pigmented siblings never produced spotted offspring in the following generations. Apparently the Black Molly males contributed to the genome of the Amazon Mollies. This paternal contribution is transmitted through the germ line and has reappeared in the soma up to now for nine generations. Multilocus DNA fingerprinting failed to reveal any differences between spotted and black fish, their *P. formosa*-like pigmented sisters, and the original maternal

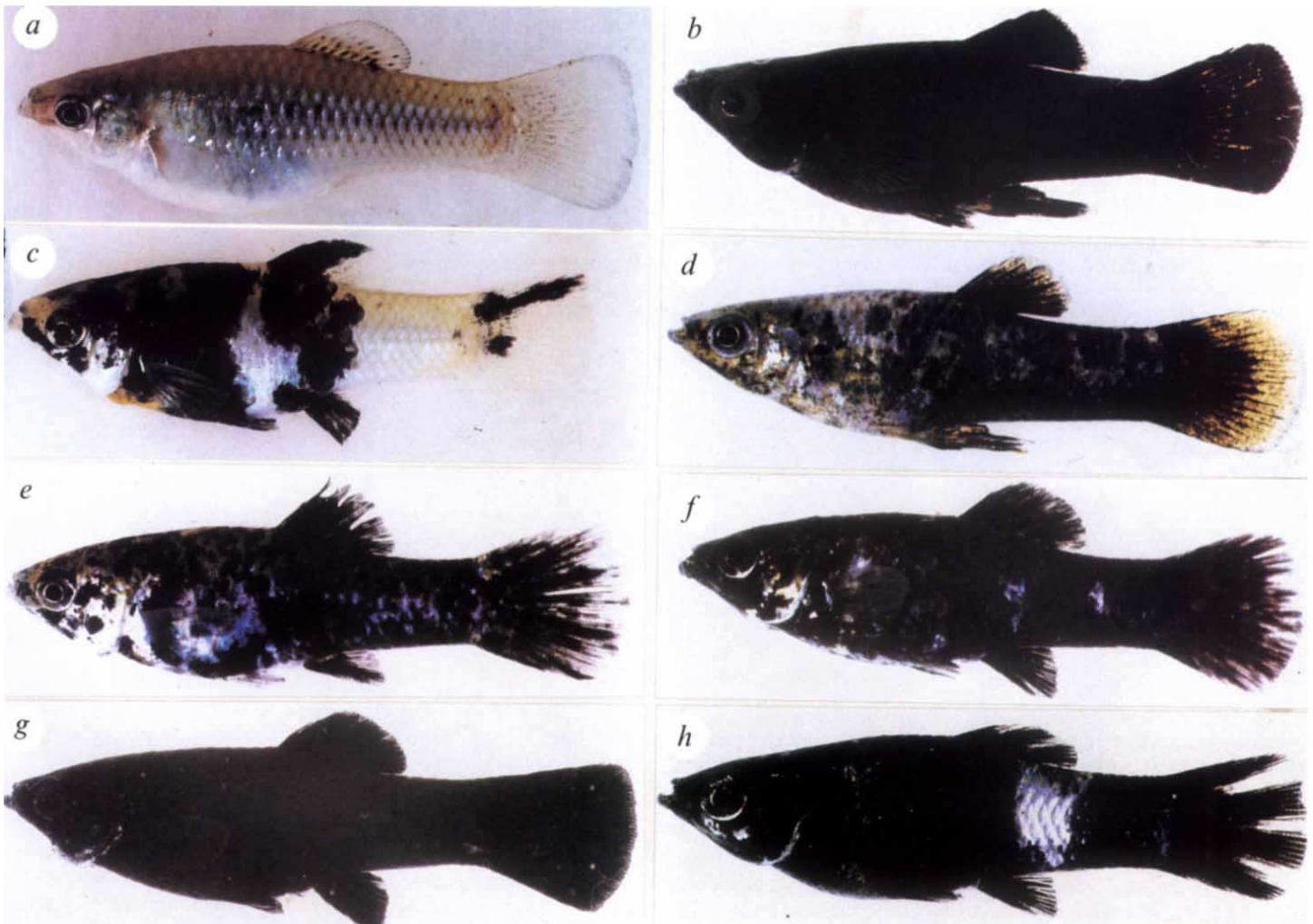
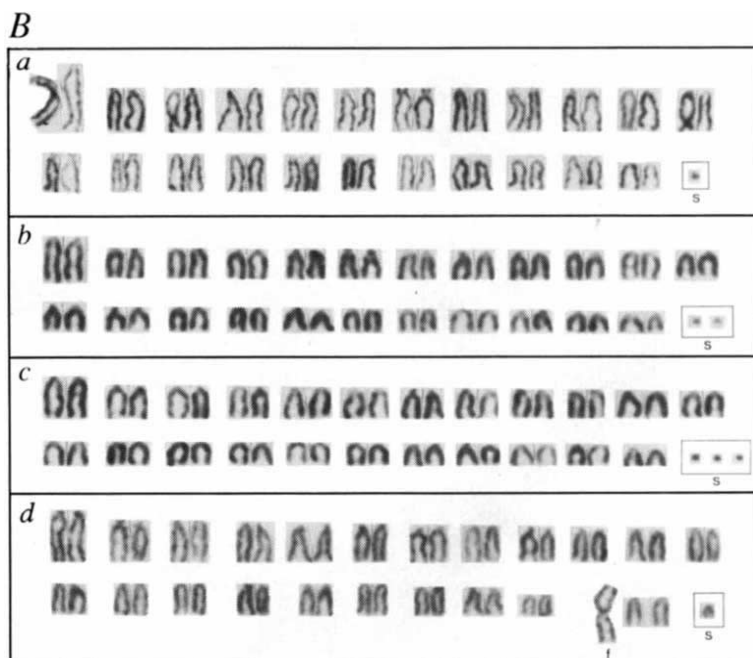
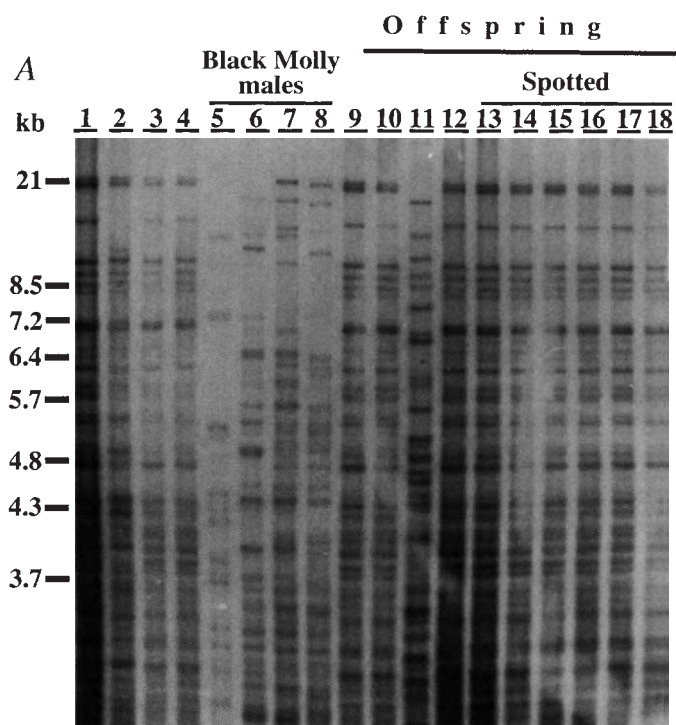


FIG. 1 Phenotypes of parental fish and unusually pigmented Amazon Mollies. a, Wild-type-pigmented *P. formosa*; b, Black Molly male; c, spotted *P. formosa* from crossing O.M/O.P (Table 1), founder female; d, F₁ hybrid *P. mexicana* × Black Molly (note the even distribution of spots compared with the animal in c); e, spotted *P. formosa*, daughter of

founder female; f, (granddaughter of founder female from crossing 3.2/3.P (Table 1); g, black *P. formosa*; sister to fish in f; h, pigmentation mosaicism in a black *P. formosa*, great-granddaughter of founder female, exhibiting a congenital large unpigmented area where the Black Molly type pigment cells are missing.

FIG. 2 A, DNA fingerprints of parental fish and unusually pigmented offspring showing identical patterns for fish from the parental strain (lane 1), the founder animal (lanes 3 and 4, muscle and liver) of the pigmented pedigree and its offspring (daughters: lane 9, (wt) greyish, wild-type-pigmented; lane 10, heavily, dark spotted; granddaughters: lane 12, homogeneously black; lanes 13, 14, spotted; great-granddaughters: lanes 15–18, spotted). Lane 2, individual from *P. formosa* strain II, which is a sub-strain separated from strain I several years ago⁹; note subtle differences in the DNA fingerprint patterns between strains I and II; lanes 5–8, Black Molly males derived from a randomly inbred laboratory stock that were used for the breeding experiments; lane 11, *P. mexicana*, male. B, Karyotypes of *P. formosa* containing supernumerary microchromosomes, s (framed). a, Wild-type-pigmented sibling; b, spotted animal; c, completely black specimen; d, feral *P. formosa* from natural habitat. In addition a centric fusion chromosome, f, with a large supernumerary chromosome, s, is present.

METHODS. A, DNA was extracted from pooled organs of individual fish, 5 µg was digested with *Hinf*I, separated on 0.8% agarose gels, and hybridized to simple repeat oligonucleotides as described^{9,29}. Gels were hybridized with ³²P-labelled probes (GGAT)₄, (CA)₈, (GATA)₄, (GAA)₆ or (GACA)₄ (not shown). B, Metaphase chromosomes were obtained after colchicine treatment of the animals from gills, kidney and spleen, according to ref. 29 and stained with Giemsa solution.



P. formosa clone (Fig. 2A). Depending on the simple repeat sequence content and the organization of the respective part of the genome, the sensitivity of this method for detection of foreign sequences would be ~5%. The paternal contribution therefore was expected to be less than 10⁸ base pairs. This explains the general *P. formosa* phenotype and the otherwise clonal appearance of the black and spotted Amazon Mollies evident from morphometric and meristic data, lack of homograft rejection, sexual behaviour, all-female sex and gynogenesis.

The structural equivalent of the paternal contribution was, however, detected by cytogenetic analysis. In addition to the normal chromosome complement (2n=46), all fish analysed from the pigmented *P. formosa* pedigree had tiny, centromere-containing supernumerary chromosomes, so-called microchromosomes (Fig. 2B, a–c). Such structures were not found in

our laboratory stocks of *P. formosa* and Black Molly. Light-pigmented siblings had one microchromosome, metaphases from spotted animals contained predominantly two microchromosomes, and from totally black individuals had predominantly three (Table 2). Thus the presence of the genes for black-body pigmentation is strongly correlated with the presence of one or two of these microchromosomes ($P < 0.0001$ by Fisher exact test). The microchromosomes did not become a stable component of the genome, because about half of the offspring of pigmented females had fewer microchromosomes than their mother. In consecutive broods the proportion of unpigmented offspring may increase (30–100%). This demonstrates some germline instability for the 2- and 3-microchromosome situation. But no animal from this pedigree was found to have lost all microchromosomes. Very rarely a somatic chromosomal instability is obvi-

TABLE 1 Inheritance of pigmentation

Pedigree	Female	Male	Offspring		
			Wild-type	Spotted	Black
0.M/0.P*	<i>P. formosa</i>	Black Molly	>100	1	0
1.0/1.P	Spotted from 0.M/0.P	Black Molly	21	2	0
2.1/2.P	Spotted from 1.0/1.P	Black Molly	0	4	2
3.1/3.P	Black from 2.1/2.P	<i>P. mexicana</i> †	5	9	5
3.2/3.P	Spotted from 2.1/2.P	<i>P. mexicana</i> †	27	8	0
504	Spotted from 3.1/3.P	<i>P. mexicana</i>	2	13	1
510	Spotted from 3.2/3.P	<i>P. mexicana</i>	44	16	1
511/1	Black from 3.1/3.P	<i>P. mexicana</i>	13	11	1
511/2	Spotted from 3.1/3.P	<i>P. mexicana</i>	32	14	0
543	Black from 511/1	<i>P. mexicana</i>	68	19	9
523	Wild type from 3.1/3.P	Black Molly	27	0	0
572	Wild type from 543	Black Molly	33	0	0
524	Spotted from 3.1/3.P	<i>P. mexicana</i>	16	17	0
533/537/536	Spotted‡	<i>P. mexicana</i>	175	186	5
HH510/504	Spotted‡	<i>P. mexicana</i>	100	77	1
Total§			526	430	25

* 0.M/0.P was a routine mating of a female from our *P. formosa* strain I (ref. 9).

† *P. mexicana* is wild-type-pigmented (olivaceous) like *P. formosa*; the same male was used for both crossings.

‡ Selection breeding stocks derived from pedigrees 504 and 510.

§ Total number of the three phenotypes in the offspring of spotted and black females in all pedigrees (this table, and data not shown).

TABLE 2 Microchromosomes in offspring of spotted and black Amazon Mollies

Phenotype	No. of microchromosomes per individual*		
	1	2	3
Wild type	10	0	0
Spotted	0	16	0
Black	0	0	6

* 30 metaphases of each specimen were analysed. In 80% of the metaphases analysed, the number of microchromosomes shown was observed; the quality of the remaining metaphases prevented accurate counting of microchromosomes.

There is another consideration besides the possible advantage gain by an ameiotic species from the acquisition of foreign genes. Host males invest much time and energy in insemination of *P. formosa* without the reward of reproductive success. This situation has been termed 'sexual parasitism'²⁵. Host males may balance their costs as a result of mate copying by conspecific females²⁶. But if a male is able to introduce some of his genes via microchromosomes into *P. formosa*, he would gain a considerable input of his own genes into the following generations as a result of the clonal mode of inheritance in *P. formosa*.

Our findings can explain how *P. formosa* and other potentially non-recombining organisms survive longer^{9,14} than predicted by models based on the accumulation of deleterious mutations^{7,8}. Contrary to the usual assumption^{2,27}, gynogenesis may not be an imperfect parthenogenesis, but a well adapted reproductive mode that combines advantages of sexual and asexual reproduction. This supports the idea²⁸ that ameiotic organisms are so rare not because sexuality is so advantageous but because the circumstances under which asexuality may arise are exceptional. □

ous from black individuals that have congenital large areas where macromelanophores are missing (Fig. 1h).

The incorporation of genetic material through microchromosomes appears to be a mechanism that may also be a relevant factor in natural populations of *P. formosa*. We found one microchromosome per metaphase in three out of twelve individuals from seven different localities, one being remarkably large (Fig. 2B, d). These microchromosomes are inherited, as demonstrated by all five daughters and the three granddaughters analysed from a wild *P. formosa*. Also, the larger supernumerary chromosome was detected in the single daughter and the three granddaughters of the wild female. These and other data²² on two of six wild Amazon Mollies having a microchromosome suggest that microchromosomes are reasonably frequent and heritable under natural conditions.

The selective pressure exerted by the experimenter for the pigmentation phenotype, leading to retention of specific microchromosomes with the macromelanophore genes, could be equal to selection in the natural habitat for any advantageous trait conferring either additional fitness or compensation for malfunctioning genes. Optimally, beneficial host species genes may become stable integrates of the Amazon Molly genome.

Even a small amount of conventional genetic recombination may be sufficient to halt population extinction owing to stochastic accumulation of deleterious mutations (Muller's ratchet)²³. It is unknown if sporadic introduction of small pieces of genetic material from the sexual host species into the parthenoform is comparably effective. The speed with which the ratchet will produce extinction of an asexual lineage depends on factors such as effective population size and mode of interaction between the mutations concerned^{23,24}. The phenomenon that we describe may certainly influence the operation of the ratchet, but to what extent remains to be investigated.

Received 18 August; accepted 7 November 1994.

- Williams, G. C. *Sex and Evolution* (Princeton Univ. Press, New Jersey, 1975).
- Maynard Smith, J. *The evolution of Sex* (Cambridge Univ. Press, Massachusetts, 1978).
- Bell, G. *The Masterpiece of Nature* (Univ. California Press, San Francisco, 1982).
- Kondrashov, A. S. *J. Hered.* **84**, 372–387 (1993).
- Charlesworth, B. *Trends Ecol. Evol.* **4**, 64–267 (1989).
- Muller, H. J. *Mutat. Res.* **1**, 2–9 (1964).
- Lynch, M. & Gabriel, W. *Evolution* **44**, 1725–1737 (1990).
- Gabriel, W., Lynch, M. & Bürger, R. *Evolution* **47**, 1744–1757 (1993).
- Schartl, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8759–8763 (1991).
- Avise, J. C., Trexler, J. C., Travis, J. & Nelson, W. S. *Evolution* **45**, 1530–1533 (1991).
- Spolsky, C. M., Phillips, C. A. & Uzzell, T. *Nature* **356**, 706–708 (1992).
- Hedges, S. B., Bogart, J. P. & Maxson, L. R. *Nature* **356**, 708–710 (1992).
- Quattro, J. M., Avise, J. C. & Vrijenhoek, R. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 348–352 (1992).
- Schartl, M., Wilde, B., Schlupp, I. & Parzefall, J. *Evolution* (in the press).
- Hubbs, C. L. & Hubbs, L. C. *Science* **76**, 628–630 (1932).
- Turner, B. J. in *Mechanisms of Speciation* (ed. Barigozzi, C.) 265–305 (Liss, New York, 1982).
- Rasch, E. M., Monaco, P. J. & Balsano, J. S. *Histochemistry* **73**, 515–533 (1982).
- Schröder, J. H. *Zool. Beitr.* **10**, 369–463 (1964).
- Balsano, J. S., Rasch, E. M. & Monaco, P. J. in *Ecology and Evolution of Livebearing Fishes (Poeciliidae)* (eds Meffe, G. K. & Snelson, F. F. Jr) 277–297 (Prentice Hall, Englewood Cliffs, NJ, 1989).
- Kallman, K. D. *Genetics* **50**, 260–262 (1964).
- Rasch, E. M. & Balsano, J. S. in *Evolution and Ecology of Unisexual Vertebrates* (eds Dawley, R. M. & Bogart, J. P.) Bulletin 466, 252–267 (New York State Museum, Albany, New York, 1989).
- Sola, L., Rossi, A. R., Bressanello, S., Rasch, E. M. & Monaco, P. J. *Cytogenet. Cell Genet.* **63**, 189–191 (1993).
- Charlesworth, D., Morgan, M. T. & Charlesworth, B. *Genet. Res.* **61**, 39–56 (1993).
- Kondrashov, A. S. *Genetics* **137**, 311–318 (1994).
- Hubbs, C. *Interactions between a Bisexual Fish Species and its Gynogenetic Sexual Parasite* Bulletin 8 (Texas Memorial Museum, Austin, 1964).
- Schlupp, I., Marler, C. & Ryan, M. J. *Science* **263**, 373–374 (1994).
- Vrijenhoek, R. C. in *Speciation and its Consequences* (eds Otte, B. & Endler, J. A.) 386–400 (Sinauer, Sunderland, 1989).
- Vrijenhoek, R. C. in *Evolution and Ecology of Unisexual Vertebrates* (eds Dawley, R. M. & Bogart, J. P.) Bulletin 466, 24–31 (New York State Museum, Albany, New York, 1989).
- Nanda, I. et al. *J. molec. Evol.* **30**, 456–462 (1990).

ACKNOWLEDGEMENTS. We thank M. J. Ryan for critically reading the manuscript and T. Grimm for statistical analysis. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (J.P. and M. Schartl), a fellowship from the HSP II program from the Deutscher Akademischer Austauschdienst (I.S.), and a grant from the Commission of the European Communities to M. Schmid. Parts of the stocks used in this study were collected under permit number 242.2190276/36299, for which we thank the Mexican government.

Increased male fertility in *Tribolium confusum* beetles after infection with the intracellular parasite *Wolbachia*

Michael J. Wade & Nancy W. Chang

Department of Ecology and Evolution, The University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

THE cytoplasmically inherited microorganism *Wolbachia pipientis* behaves like a sexually selected trait in its host, the flour beetle *Tribolium confusum*, enhancing male fertility at the expense of female fecundity. Here we show that infected females have fewer offspring than uninfected females but infected males have a large fertility advantage over uninfected males within multiply-inseminated infected or uninfected females. The male fertility effect accelerates the spread of the *Wolbachia* through the host population and expands the initial opportunity for hitch-hiking of host nuclear genes. Sperm competition in a host, mediated by endosymbionts, has not been previously described.

Wolbachia pipientis is a member of the Rickettsiaceae (Gram-negative bacteria). These obligate intracellular pathogens of eukaryotes¹ are found in invertebrate hosts but cause several vertebrate diseases, including typhus, Rocky Mountain spotted fever and Q fever in humans. The endosymbiotic ancestor of the mitochondrion is believed to have been a member of this group². A 16S ribosomal sequence phylogeny³⁻⁵ shows that the mammalian pathogens *Anaplasma marginale*, *Cowdria ruminantium* and *Ehrlichia* are among the closest relatives of *W. pipientis*. The pathogenic effects of these endosymbionts in mammals are better understood than the fitness effects they have on their invertebrate hosts⁶⁻¹⁴.

The genus *Wolbachia*^{6,15-18} occurs in over 90 arthropod species, including five orders of insects, an isopod and a spider mite. In most cases, infected males produce reduced numbers of progeny with uninfected females and *Wolbachia* species have been called cytoplasmic incompatibility microorganisms (CIMs). Progeny reduction varies from complete (none produced) in some hosts, to partial or no reduction in others¹⁹. In *Nasonia*¹³, *Drosophila*²⁰ and *Aedes*^{21,22}, bidirectional incompatibility between populations or species has been observed: the elimination of *W. pipientis* from both host populations with antibiotic medium restores interfertility. This suggests a possible role in speciation^{13,23}.

In the flour beetle host, *Tribolium confusum*, complete incompatibility occurs between infected (I) males and uninfected (U) females^{6,8}. Rarely, a few but sterile offspring are produced in incompatible crosses by older U-females (Table 1). The incompatibility may result from the aborted syngamy of the maternal (U) and paternal (I) genomes^{13,16,18,24}, because CIMs cause inviability in diploids but viable haploid male offspring in haplodiploids¹¹.

Stevens⁸ suggested that *W. pipientis* may manipulate the behaviour of its host to its own advantage and at the expense of host fitness¹⁸. *Tribolium* flour beetles mate several times a day²⁵ and U-females could avoid laying inviable eggs by discriminating against I-males as mates or against their stored sperm post-copulation. To test this, we mated single U-females in 8 g of standard medium²⁶ to many males, varying the number of I- and U-males among treatments (Table 1). In all crosses, we used a stock collected in Zagreb, Croatia (formerly Yugoslavia), in 1979, which harbours *W. pipientis*⁷. We placed several hundred adults from this I-stock on medium containing antibiotics^{7,9} and reared the offspring (>500) to create an uninfected (U) subpopulation. Virgin males and females from these stocks were used in all experiments. We varied female age

and the time available for reproduction. We let one set of U-females reproduce for the first 18 days of adult life post-pupal eclosion and then transferred them with mates to 8 g of fresh medium for an additional 18 days of reproduction before discarding the parents (Table 1; columns (1) and (2), respectively). In a second set, U-females reproduced in 8 g of medium with one or more mates for the first 53 days of adult life before we transferred them for an additional 53 days of reproduction (Table 1; columns (3) and (4)). We counted all offspring maturing to adults (Table 1).

If U-females discriminate against I-males or their sperm to avoid producing inviable eggs, then offspring numbers should be higher than expected under random mating⁸. In fact, we observed the opposite: the number of offspring produced by multiply mated U-females is lower than expected in all but two cases (Table 1). The reduction in offspring is especially severe in the early reproductive life of females (0–18 days; column (1), Table 1) but diminishes with female age (Table 1, per cent reduction in offspring: (1) > (2) and (3) > (4)). To rule out non-random mating, we set up a control with 120 virgin adult beetles: 30 I-males and 30 I-females (homozygous for a wild-type body-colour marker; see below) and 30 U-males and 30 U-females (homozygous for a black body-colour marker). We observed 56 copulations but did not find deviations from random mating ($\chi^2 = 0.22$, d.f. = 1, $P = 0.64$) nor any tendency of I-males to mate first. We infer that the large reduction in offspring numbers is owing to a fertility advantage of I-males in post-copulatory sperm competition with U-males. (A post-copulatory advantage to conspecific over heterospecific sperm occurs in matings between other species in this genus^{26,27}.)

We also mated single I-females in 8 g of standard medium to varying numbers of I- and U-males (Table 2). To distinguish paternity, all I-females and I-males were $+/+$ homozygotes for a red body-colour allele and the U-males were b/b homozygotes for a black body-colour allele at the same locus. Heterozygous offspring ($+/b$) are intermediate in body colour between red and black. We tested the marker alleles for fitness effects (Fig. 1) and found a slight advantage for the b allele at intermediate frequencies. The stock was homozygous for the black allele but the wild-type red body-colour allele arose during laboratory husbandry and segregates in the stock. For this reason, the fertility advantage we report next for wild-type I-males in mating competition with homozygous black U-males may be an underestimate.

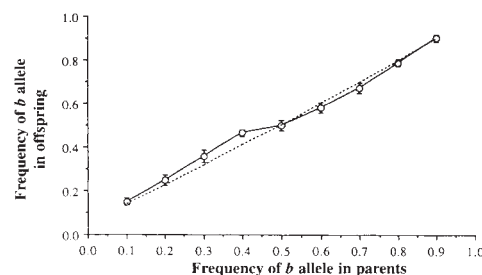


FIG. 1 The relationship between b allele frequency in the parents and offspring. Replicate groups of 10 males and 10 females were set up at 11 different initial frequencies of the b allele by changing the frequency of pairs of homozygous b/b and $+/+$ (wild-type) adults. Adults of both genotypes carried *W. pipientis*. We counted all offspring produced by each group of twenty I-adults and scored the genotypes. The observed mean frequencies (circles) of the b allele in the offspring are shown with one standard error. The dotted line represents the neutral expectation of no selection at this locus. Groups of twenty homozygous $+/+$ I-adults produce 10% more offspring than groups of 20 b/b I-adults (153.6 versus 139.5, respectively) but the difference is not statistically significant. Allowing for this fecundity difference at intermediate allele frequencies, there appears to be a slight fitness advantage to the black allele which results in an excess production of heterozygotes.

TABLE 1 Offspring numbers from single uninfected females mated with males infected or uninfected with *Wolbachia pipientis*

Number of males		(1) 0–18 days		(2) 19–36 days		Per cent reduction in offspring		Exp.
I	C	Mean (X)	s.e.	Mean (X)	s.e.	(1)	(2)	
0	1	111.4	3.8	99.8	7.5	—	—	—
1	0	0.0	—	0.0	—	100.0	100.0	100.0
0	2	93.8	4.0	108.4	3.9	—	—	—
1	1	9.8	3.5	44.4	10.7	89.6***	59.0	50.0
2	0	0.0	—	0.0	—	100.0	100.0	100.0
1	3	25.8	7.3	55.0	13.8	72.5***	50.7**	25.0
2	2	12.0	2.5	45.6	8.1	87.2***	57.9	50.0
3	1	13.4	3.5	32.4	6.2	85.7*	70.1	75.0
1	5	42.1	10.5	51.0	12.0	55.1***	53.0***	16.7
3	3	10.0	1.4	35.6	4.6	89.3***	66.2**	50.0
5	1	2.8	0.6	15.2	3.6	97.0***	86.0	83.3

		(3) 0–53 days		(4) 54–106 days		(3)	(4)	Exp.
I	C	Mean (X)	s.e.	Mean (X)	s.e.			
0	1	96.8	19.1	143.7	16.7	—	—	—
1	0	0.0	0.0	4.7	3.3	100.0	96.7	100.0
0	2	135.2	29.3	146.2	15.7	—	—	—
1	1	56.8	12.2	73.4	15.4	58.0	50.0	50.0
2	0	0.0	0.0	1.0	2.2	100.0	99.3	100.0
0	4	138.5	15.0	170.4	8.0	—	—	—
1	3	69.0	30.0	86.7	22.4	50.2***	49.1***	25.0
2	2	17.7	7.3	35.8	11.3	87.2***	79.0***	50.0
3	1	1.2	1.3	9.8	7.4	99.1***	94.4***	75.0
4	0	0.0	0.0	0.2	0.2	100.0	99.9	100.0

χ^2 tests of observed mean relative to expected, 1 d.f.; *** $P < 0.001$, ** $P < 0.010$, * $P < 0.038$; I, infected; C, uninfected; Exp, the expected fraction of offspring assuming random mating and equal fertility of I- and C-males.

TABLE 2 Offspring numbers of I-males competing with C-males to sire offspring with single I-females

Number of males		(1) 0–18 days		(2) 19–36 days		Per cent of total offspring		Exp.
I	C	Mean (X)	s.e.	Mean (X)	s.e.	(1)	(2)	
0	1	124.6	8.9	126.2	69.0	—	—	—
1	0	124.6	9.5	115.0	—	100.0	100.0	100.0
0	2	100.0	13.2	102.7	16.0	—	—	—
1	1	13.8	5.4	73.5	16.8	15.0***	59.3	50.0
2	0	127.5	18.4	119.5	13.4	100.0	100.0	100.0
1	3	34.6	21.6	49.0	13.2	39.0**	50.4***	25.0
2	2	59.2	11.9	65.6	8.1	64.9*	56.4	50.0
3	1	59.5	16.3	83.5	12.9	66.5	80.3	75.0
1	5	64.0	13.1	55.6	22.6	65.0***	55.4***	16.7
3	3	50.4	7.5	45.2	5.0	65.1*	53.8	50.0
5	1	94.8	20.3	93.5	21.7	96.9	85.2	83.3

		(3) 0–53 days		(4) 54–106 days		(3)	(4)	Exp.
I	C	Mean (X)	s.e.	Mean (X)	s.e.			
0	1	37.0	7.1	93.3	15.9	—	—	—
1	0	49.0	5.0	104.5	12.1	100.0	100.0	100.0
0	2	38.7	9.2	119.7	28.3	—	—	—
1	1	24.3	3.5	54.7	15.0	56.1	56.8**	50.0
2	0	65.0	13.1	101.5	15.5	100.0	100.0	100.0
1	2	19.7	4.4	56.7	29.0	49.3***	58.8***	33.3
2	2	21.7	6.0	33.7	3.9	53.3	47.0	50.0
1	4	14.0	8.8	28.0	11.4	33.6***	35.8***	25.5

When no I males are present, offspring are sired by C-males. χ^2 test of observed mean relative to expected, 1 d.f., *** $P < 0.005$, ** $P < 0.015$, * $P < 0.04$.

When an I-female is multiply inseminated by I- and U-males, the numbers of offspring sired by I-males tends to equal or exceed expectation, with few exceptions (Table 2). In some instances, a single I-male reduces the offspring sired by two or three U-males by 15–25%. The pattern of paternity does not change with age of the I-females as it did with U-females (Table 2, per cent of total offspring: (1) versus (2) and (3) versus (4)).

An I-male fertility effect accelerates the spread of *W. pipientis* through the host population (Table 3). The change in frequency of infected hosts per generation is $(p_1' - p_1) = w_1(p^2q)/(1 - pqw_1)$. When infected hosts are rare ($p_1 \sim 0$), the ratio of the rate with the I-male fertility advantage ($w_1 > 1$) to that without ($w_1 = 1$) is equal to w_1 : the acceleration is directly proportional to the relative fertility advantage of I-males over U-males. We observed

TABLE 3 Spread of *W. pipientis* and hitch-hiking of a host nuclear gene

Males	Females	+/+	Offspring +/b	b/b
q +/+, U p b/b, I	b/b, I	—	qw _u	pw _i
q +/+, U p b/b, I	+ /+, U	qw _u	—	—
$p_i' = p_i / (1 - pqw_i)$ $p_b' = (p_b)(2 - q + w_u) / (1 - pqw_i)$				

Infected (I) hosts, homozygous *b/b* and in frequency *p*, are invading an uninfected (U) population, homozygous for an alternative allele, *+/+*, in frequency *q* = (1 - *p*). The relative fertility advantage to I-males is *w_i* > 1 and the relative disadvantage of U-males is *w_u* < 1. Note that (*pw_i* + *qw_u*) = 1. We let females mate essentially an infinite number of times and use primes to signify the frequencies of infected hosts (*p_i'*) and *b*-alleles (*p_b'*) in the offspring.

ratios of *w_u*/*w_i* between 0.35 and 0.50 (Tables 1 and 2) which could cause an accelerated spread comparable to that observed in the very rapid spread of *W. pipientis* through northern California populations of *Drosophila simulans*¹¹, where it was attributed to high migration rates of 0.22 per population per generation.

Despite the deleterious effect on U-females of mating with I-males, there is no tendency for females to discriminate against these males or their sperm. The effects of *W. pipientis* on host fitness in *T. confusum* are similar to those of a sexually selected trait that enhances male fertility at the expense of other fitness components, like viability, in males or females²⁸. *W. pipientis* has opposing and sex-specific effects on host fitness indicative

of a complex relationship between host and microorganism¹⁸. Studies of host-disease dynamics in model systems like *Tribolium* assist our understanding of the evolutionary genetics of the Rickettsiaceae, a taxonomically important group of vertebrate pathogens and arthropod endosymbionts. □

Received 10 June; accepted 8 November 1994.

1. Woese, C. R. *Microbiol. Rev.* **51**, 221-271 (1987).
2. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4443-4447 (1985).
3. O'Neill, S. L., Giordano, R., Colbert, A. M. E., Karr, T. L. & Robertson, H. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2649-2702 (1992).
4. Breeuwer, J. A. J. et al. *Insect molec. Biol.* **1**, 25-36 (1992).
5. Stouthamer, R., Breeuwer, J. A. J., Luck, R. F., Werren, J. H. *Nature* **361**, 66-68 (1993).
6. Wade, M. J. & Stevens, L. *Science* **227**, 527-528 (1985).
7. Stevens, L. & Wade, M. J. *Genetics* **124**, 367-372 (1990).
8. Stevens, L. *Anim. Behav.* **46**, 305-310 (1993).
9. Kambhampati, S., Rai, K. S. & Burgun, S. J. *Evolution* **47**, 673-677 (1993).
10. Hoffmann, A. A. & Turelli, M. *Genetics* **119**, 435-444 (1988).
11. Turelli, M. & Hoffmann, A. A. *Nature* **353**, 440-442 (1991).
12. Turelli, M., Hoffmann, A. A. & McKechnie, S. W. *Genetics* **132**, 713-723 (1992).
13. Breeuwer, J. A. J. & Werren, J. H. *Nature* **346**, 558-560 (1990).
14. Montchamp-Moreau, C., Feveur, J. & Jacques, M. *Genetics* **129**, 399-407 (1991).
15. O'Neill, S. L. & Karr, T. L. *Nature* **348**, 178-180 (1990).
16. Rousset, F. & Raymond, M. *Trends Ecol. Evol.* **6**, 54-57 (1991).
17. Wade, M. J. & Stevens, L. *J. theor. Biol.* **167**, 81-88 (1994).
18. Moran, N. & Baumann, P. *Trends Ecol. Evol.* **9**, 15-20 (1994).
19. Holden, P. R., Jones, P. & Brookfield, J. F. Y. *Genet. Res.* **62**, 23-29 (1993).
20. Rousset, F. & De Stordeur, E. *Heredity* **72**, 325-331 (1993).
21. Trips, M. J., Perrone, J. B., Reissig, M. & Parker, K. L. *J. Heredity* **72**, 313-317 (1993).
22. Subbarao, S. K. in *Recent Developments in the Genetics of Insect Disease Vectors* (eds Steiner, W. W. M., Tabachnick, W. J., Rai, K. S. & Narang, S.) 313-342 (Stipes, Champaign, Illinois).
23. Thompson, J. N. *Biol. J. Linn. Soc.* **32**, 385-393 (1987).
24. Richardson, P. M., Holmes, W. P. & Saul II, G. B. *J. Invert. Path.* **50**, 176-183 (1987).
25. Park, T. J. *exp. Zool.* **65**, 17-42 (1933).
26. Wade, M. J., Patterson, H., Chang, N. W. & Johnson, N. A. *Heredity* **72**, 163-167 (1994).
27. Robinson, T., Johnson, N. A. & Wade, M. J. *Heredity* **73**, 155-159 (1994).
28. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721-3725 (1981).

ACKNOWLEDGEMENTS. We thank N. Johnson, J. Kelly, L. Stevens, M. Antolin, J. Koehn, M. Whitlock, D. Charlesworth and C.-I. Wu for discussion and/or comments on the manuscript, and M. McNaughton for technical assistance. This work was supported by an NIH grant.

A transcription factor controlling development of peripheral sense organs in *C. elegans*

Connie Zhao & Scott W. Emmons

Department of Molecular Genetics,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, New York 10461, USA

THE basic-helix-loop-helix (bHLH) proteins constitute a class of transcription factors thought to be important in the control of cell-type determination¹. These transcription factors are believed to activate the expression of cell-type-specific genes to generate stable differentiated cell types². The expression of bHLH proteins, in turn, is regulated by spatial cues, so that switches in cell type occur in a reproducible pattern³. We report here that the *lin-32* gene of *Caenorhabditis elegans*, which encodes a bHLH protein of the *Drosophila achaete-scute* family of transcription factors, is necessary and in some cells sufficient for specification of the neuroblast cell fate. Similarity in the function and structure of the *lin-32* protein (LIN-32) to transcription factors of the *achaete-scute* gene family in *Drosophila* and vertebrates implies that this class of transcription factors functioned in a primitive ancestral form to specify neuronal cell fate, supporting the proposition that certain basic mechanisms of cell-type determination have been conserved through metazoan evolution¹.

We isolated the *lin-32(bx46)* X mutation in a screen for mutations with abnormal male tail morphology. Mapping and complementation testing showed that *bx46* was allelic to two

previous mutations, isolated in screens for, respectively, mutations with lineage defects (*e1926*) (C. J. Kenyon and E. M. Hedgecock, personal communication) and with defects in the generation of touch neurons (*u282*)⁴. All the cells known to be affected in *lin-32* mutants are neurons or support cells of peripheral sense organs. The most obvious phenotype of all three alleles is that males lack rays, peripheral sense organs specialized for sensing the hermaphrodite during copulation (Fig. 1a; Table 1, lines 2-4). In both males and hermaphrodites, *lin-32* mutations also block the generation of the Q and postdeirid neurons. In addition, the *u282* allele affects the generation of PLM touch neurons, and therefore *u282* mutant animals are touch-insensitive in the tail (Table 1, line 4)⁴. Animals mutant for the other two alleles are touch-sensitive (Table 1, lines 2 and 3). We performed lineage analysis with the *u282* allele and found that the *lin-32* mutation caused transformation of the ray and postdeirid neuroblasts into hypodermal cells (Fig. 1b, c). Similar effects were observed with the *e1926* allele (C. J. Kenyon and E. M. Hedgecock, personal communication). *lin-32(u282)* also results in transformation of the Q neuroblast into a hypodermal cell that fuses with the hypodermal syncytium covering most of the body (Fig. 1d)⁴. Together, these results indicate that the *lin-32* mutant phenotype results from the loss of several neuroblasts and the transformation of these cells into hypodermal cells.

We cloned the *lin-32* gene by transformation rescue, and isolated a *lin-32* cDNA (*c52*; 1.4 kilobases (kb)) by screening a clone bank with a 5-kb rescuing genomic fragment. *c52* contains an intron and therefore represents an unprocessed form of the *lin-32* transcript (Fig. 2, legend). Two lines of evidence confirm that *c52* represents a *lin-32* transcript: a point mutation in the coding region was identified in each of the three *lin-32* alleles (Fig. 2), and expression of *c52* from a heat-shock promoter rescued *lin-32* mutants (Fig. 2, legend). The conceptual 71-amino-acid polypeptide encoded by *c52* is likely to be the full-

TABLE 1 Effect of *lin-32* mutations on per cent expression of rays, touch sensitivity and ectopic ray papillae

	Genotypes	Ray 1	Ray 2	Ray 3	Ray 4	Ray 5	Ray 6	Ray 7	Ray 8	Ray 9	n	Touch sensitivity in the tail	m
1	wild type	100	100	100	100	100	100	100	100	100	100	Sensitive	>100
2	<i>lin-32(e1926)</i>	0	3	19	20	43	65	3	0	14	260	Sensitive	100
3	<i>lin-32(bx46)</i>	0	4	10	12	24	48	0	1	2	120	Sensitive	100
4	<i>lin-32(u282)</i>	0	0	1	1	1	1	1	0	0	155	Insensitive	100
5	<i>eT2/+</i>	46	87	100	100	100	100	100	100	100	48	ND	
6	<i>lin-32(e1926)/+</i>	100	100	100	100	100	100	100	100	100	108	Sensitive	>100
7	<i>lin-32(bx46)/+</i>	100	100	100	100	100	100	100	100	100	76	Sensitive	>100
8	<i>lin-32(u282)/+</i>	58	100	100	100	100	100	100	100	100	80	Sensitive	>100
9	<i>lin-32(e1926)/eT2</i>											Insensitive	30
10	<i>lin-32(bx46)/eT2</i>											Insensitive	42
11	<i>lin-32(u282)/eT2</i>											NA (lethal)	
12	<i>lin-32(e1926)/lin-32(u282)</i>											Sensitive	100
13	<i>lin-32(bx46)/lin-32(u282)</i>											Sensitive	100

The frequency (per cent) of appearance of a given ray in males and the touch sensitivity of males and hermaphrodites. Rays were scored by Nomarski microscopy and identified by position and morphology. *n*, Number of sides scored; expression of rays on the two sides of individual males seemed to be independent. Touch sensitivity in the tail was determined by stroking the tail with an eyelash. ND, not determined; NA, not applicable. *m*, Number of animals tested for touch sensitivity. *him-5(e1490)* was present in lines 1–4 to obtain a high frequency of males. XX males generated in a *tra-1(e1099)/III* background¹¹ have a normal complement of rays. The *bx46* allele of *lin-32* was isolated in a visual screen of male progeny of ethyl methanesulphonate(EMS)-mutagenized hermaphrodites and was shown to be an allele of *lin-32* by failure to complement *lin-32(e1926)* and *lin-32(u282)*. Strain CB3690 (*him-5(e1490); lin-32[e1926]*) was obtained from C. Kenyon; strain TU282 (*lin-32[u282]*) was obtained from *Caenorhabditis* Genetics Center, and the *him-5(e1490)* mutation was introduced (EM321). The homozygous lethal rearrangement *eT2* was generated by C. Kenyon and sent to us by D. Greenstein; it is thought to be a non-reciprocal translocation of material from the left end of the X chromosome onto chromosome I, with some X-chromosomal material lost. It fails to complement *lin-32* and genes to the left, but complements *unc-2* (A. Chisholm and D. Greenstein, personal communication). Genetic constructions were as follows: line 5: *tra-1* males were crossed to *eT2/unc-2(e55)* hermaphrodites and *F₁* non-Unc hermaphrodites were individually picked and allowed to self-fertilize. Those that segregated no Unc progeny were identified as genotypically *tra-1/+; eT2/+*. Among their *F₂* male progeny, one third were expected to be *tra-1; +/+*, and two thirds were expected to be *tra-1; eT2/+*. A total of 72 sides of such male progeny were scored; 20 sides were missing ray 1 and 6 sides were missing both ray 1 and ray 2. Because theoretically two thirds, or 48 sides, were of males genotypically *eT2/+*, the frequency of ray 1 and ray 2 loss in *eT2/+* males was estimated as 26/48 (54%) and 6/48 (13%), respectively. Lines 6–8: Males heterozygous for the three alleles of *lin-32* were *F₁* non-Unc progeny of a cross between *tra-1* males and *unc-4(e120); tra-1; lin-32; eDp6(III; f)* hermaphrodites. *eDp6(III; f)* is a free duplication covering *tra-1*; it is lost during meiosis with a frequency of 50%, resulting in generation of *tra-1* XX males. Lines 9–11: *tra-1* males were crossed with *tra-1; lin-32; eDp6* hermaphrodites, and *F₁* males (*tra-1; lin-32/+*) were mated to *unc-2(e55)/eT2* hermaphrodites. The four expected classes of non-Unc cross-progeny were discriminated by self-progeny testing. For the three *lin-32* alleles, the number of each class (*lin-32/eT2; +/eT2; lin-32/unc-2; +/unc-2*) was as follows: *e1926*, 7, 8, 9, 11; *bx46*, 8, 8, 12, 10; *u282*, 0, 15, 17, 17. All *lin-32(e1926)/eT2* and *lin-32(bx46)/eT2* hermaphrodites were touch insensitive in the tail. Lines 12 and 13: *tra-1* males were mated to *tra-1; lin-32(u282); eDp6* hermaphrodites, and *tra-1; lin-32(u282)/+* cross progeny males were mated to *unc-4(e120); tra-1; lin-32(e1926); eDp6* and to *unc-4(e120); tra-1; lin-32(bx46); eDp6* hermaphrodites. One half of the non-Unc progeny are of the desired *trans* double-heterozygote genotype. All of 200 tested non-Unc progeny were touch-sensitive.

length *lin-32* gene product. First, multiple in-frame stop codons are present in the cDNA upstream of the ATG initiation codon. Second, a sequenced genomic region of 400 nucleotides upstream of the complementary DNA contained no potential additional 5' exons. Finally, a genomic fragment containing an additional 400 base pairs (bp) 5' of the cDNA rescued a *lin-32* mutation (data not shown), suggesting that this region contains the *lin-32* promoter.

The predicted LIN-32 is a putative transcription factor containing a bHLH domain⁵ (Fig. 2). The LIN-32 bHLH region is similar to those of the *Drosophila achaete-scute* genes and their homologues in vertebrates. The most similar protein is *Drosophila* atonal (63% amino-acid identity in the bHLH domain), which functions in expression of the neuroblast cell fate by precursor cells of chordotonal organs and by the R8 photoreceptor of the eye^{6,7}. LIN-32 has a single amino acid (M) on the N-terminal side of the bHLH domain, and 18 amino acids (SQILKQDSKNENLKS KSG) on the C-terminal side, making it the smallest member of the family. It does not have the C-terminal acidic domain present in some other bHLH proteins⁸; the acidic domain is not required for certain bHLH proteins to function as transcription factors^{9,10}.

We used genetics to explore the role of the putative LIN-32 transcription factor in determination of the neuroblast cell fate. We found *lin-32* is likely to be a key component of the switch to this developmental fate. The three mutant alleles seem to lower the activity of *lin-32*; this results in a decreased probability of a cell assuming the neuroblast cell fate. Ectopic expression of *lin-32*, on the other hand, promotes the neuroblast cell fate. These results indicate that regulation of the level of *lin-32* activity

is a central step in the generation of neuroblasts. The three alleles of *lin-32* form an allelic series (Table 1). *u282* is the strongest allele, causing 99% ray loss and touch insensitivity in the tail. *bx46* and *e1926* cause respectively 88 and 81% ray loss and have normal touch-sensitivity. *u282* is semidominant for loss of ray 1, whereas *e1926* and *bx46* are recessive (Table 1, lines 6–8). Because *C. elegans* males are hemizygous for the X chromosome and *lin-32* is sex-linked, in these and the following experiments in which *lin-32* heterozygotes are examined, we generated males diploid for the X chromosome by taking advantage of a mutation in the sex-transformer gene *tra-1* (ref. 11). When two X chromosomes are present, dosage compensation results in the reduction of expression of X-linked genes to the level found when only a single X is present¹². This allowed us to show directly that a decreased level of wild-type *lin-32* gene activity resulted in the loss of the neuroblast cell fate, as follows. In males carrying one wild-type X chromosome and one translocation chromosome containing a deficiency covering the *lin-32* locus (*eT2*), ray 1 and ray 2 were lost with frequencies of 54 and 23%, respectively (Table 1, line 5). Because a similar frequency of ray 1 loss was found in *u282/+* males, *u282* seems to cause a strong loss of *lin-32* gene activity. For all three alleles, the phenotype of the deficiency heterozygote is more severe than the homozygote, and hence each is likely to be a hypomorph: *u282/eT2* is lethal (Table 1, line 11); *bx46/eT2* and *e1926/eT2* are viable but touch-insensitive in the tail (Table 1, lines 9 and 10). As *bx46/u282* and *e1926/u282* are both touch-sensitive (Table 1, lines 12 and 13), *u282* is not equivalent to the deficiency and is probably not a null. These data are interpreted to indicate that the level of *lin-32* gene activity decreases progressively in the

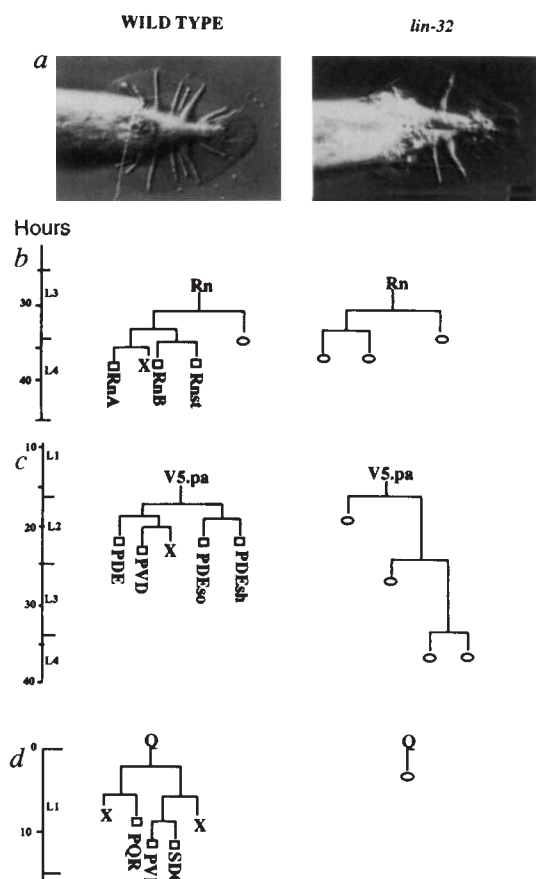
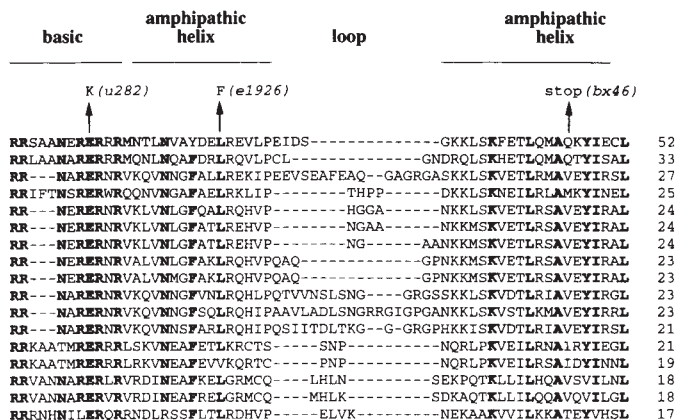


FIG. 1 *lin-32* mutant phenotype. a, Photomicrographs (Nomarski optics) of wild-type and *lin-32(bx46)* mutant adult male tails, illustrating the loss of most rays characteristic of all three *lin-32* alleles. Ventral view; anterior to the left, magnification bar, 10 μ m. b–d, Lineage alterations in *lin-32(u282)* mutants. Wild-type lineages are derived from ref. 17. b, Ray sublineages (four sides); c, postdeirid lineage (four sides); and d, Q lineage⁴. The vertical axis represents hours since hatching at 20 °C; L1–L4 represent four larval stages; the horizontal axis represents the anteroposterior order of the cells. Symbols: □, neuron or support cell; ○, hypodermal cell; X, programmed cell death. Methods for following cell lineages by Nomarski microscopy in living animals are described in ref. 17. Hypodermal cells were distinguished from neuronal cells by having larger nuclei with a flat appearance and distinct nucleoli, whereas neuronal nuclei are smaller and have a granular appearance and often indistinct nucleoli¹⁷. The hypodermal identity of Rn.a daughters in *lin-32* mutants was confirmed by staining males with MH27, an antibody that stains adherens junctions between epidermal cells in *C. elegans*, following methods described in ref. 18. Such staining showed the generation, in *lin-32* mutant males, of an extra set of epidermal cells, as expected from the lineage analysis (data not shown).

FIG. 2 Comparison of the bHLH domain of LIN-32 with representative bHLH domains. The bHLH domain of LIN-32 is aligned with bHLH domains from *Drosophila* atonal⁹ and asense proteins¹⁹, human stem-cell leukaemia protein (SCL-hu)²⁰, two mammalian achaete-scute homologues (MASH 1 and 2)²¹, three *Xenopus* achaete-scute homologues (ASH 1a, 3a and 3b)⁸, *Drosophila* lethal of scute (l'sc)²², achaete and scute, mouse MyoD²³, *C. elegans* MyoD homologue (Ce MyoD)²⁴, human E12 and E47⁵, and human N-Myc²⁵. The number of identical residues in LIN-32 and other bHLH proteins is shown on the right; dashes indicate gaps in the alignment. The 16 most conserved residues among bHLH proteins are shown in bold. Residues altered in the three *lin-32* mutants are indicated by vertical arrows. Consistent with the molecular nature of these mutations, *bx46*, but not *e1926* or *u282*, was found to be suppressed by *sup-5(e1464sd)*, an amber suppressor (data not shown).

METHODS. *lin-32* is at X chromosomal genetic map position –14.55. In order to identify the corresponding position on the physical map, a closely linked TcI DNA polymorphism in the Bergerac BO strain, *bxP5*, was identified using the general approach described in ref. 26. *bxP5* mapped to position –14.12 between *lin-32* and *unc-2*. DNA flanking the TcI element in *bxP5* hybridized to three overlapping YACS (Y38B11, Y46A5, and Y37G4) and two cosmids (C04C8 and R07D2). A second physical reference point was provided by the left endpoint of duplication *mnDp33(X; IV)*, which maps genetically to the right of *unc-2* and was placed on the physical map (YAC Y5E3)²⁷. Candidate cosmids were tested for complementation rescue of *lin-32(u282)*. Transgenic nematodes were generated as described²⁸; plasmid pRF4, bearing the dominant marker *rol-6(su1006)*, was coinjected to identify successfully transformed animals²⁸. Cosmid ZC391 resulted in restoration of rays in 90% of Rol males. A rescuing 5.0-kb *KpnI*–*Bam*HI fragment from ZC391 was identified and subcloned (EM#225). EM#225 was used to screen a cDNA library of mixed stage poly(A)⁺ RNAs²⁹. Two identical and possibly not independent clones were obtained from



2×10^6 plaques; one was studied further, c52. The ability of c52 to rescue the *lin-32(u282)* mutation was tested by ligation of the 1.4-kb cDNA to vector pPD49.78 at the *SacI* and *Apal* sites (A. Fire, S. White-Harrison and D. Dixon, personal communication). pPD49.78 contains the *C. elegans* *hsp16-2* promoter for expression of the insert. The recombinant plasmid (EM#226) was introduced as an extrachromosomal array into EM321 together with pRF4. After heat shock at 30° for 3 h during the L3 larval stage, 61% of sides ($n=116$) of Rol males had at least one ray, and overall 14% of rays were present. Control values for per cent of sides with at least one ray, and per cent of rays present were, respectively, 4 and 1% in EM321 without the transgene or heat shock ($n=155$); 6 and 1% in EM321 with the transgene without heat shock ($n=69$); 6 and 1% in EM321 without the transgene with heat shock ($n=80$). Sequence analysis suggested that c52 contained an intron of 384 nucleotides within the first helix of the bHLH domain. This was confirmed by PCR amplification of reverse-transcribed RNA, which yielded the spliced product (data not shown). Sequences of genomic DNA and cDNA are available from GenBank (accession number U15418).

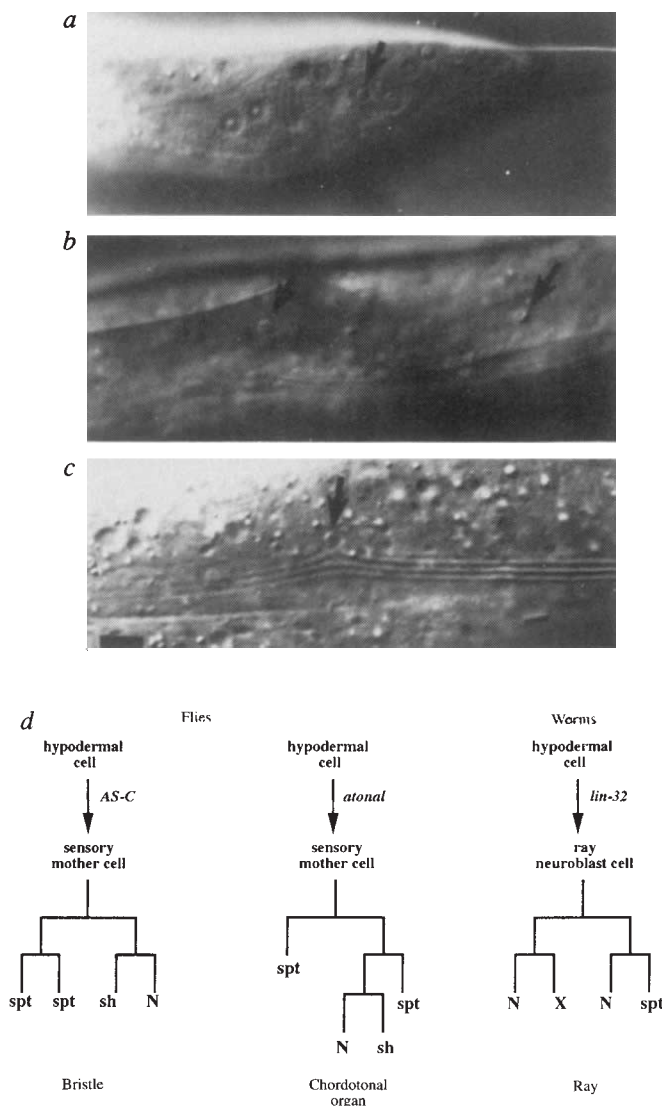


FIG. 3 a–c, Nomarski photomicrographs showing that expression of *lin-32* under the control of a heat-shock promoter induces the generation of ectopic ray papillae. Magnification bar indicates 10 μ m. Identical ray papillae with characteristic 'ring and dot' morphology are seen, a, in the tail of a wild-type L4 larval male, before morphogenesis of rays; b, in the lateral epidermis of a *lin-22(mu2)* mutant male¹³; c, in the lateral epidermis of a male carrying an integrated *hsp-lin-32* transgene, *bxln1*, after heat shock. The combined frequency of ectopic ray papillae (number of ectopic papillae scored/number of sides examined) in males and hermaphrodites of various genotypes after heat-shock treatment during the optimal interval was: *him-5(e1490)*; *bxln1*, 25/50 (50%); *him-5(e1490)*; *lin-32(u282)*; *bxln1*, 2/52 (4%); *him-5(e1490)*, 0/300 (0%). In the absence of heat shock, *him-5(e1490)*; *bxln1* had no ectopic papillae (number of sides examined, 200). Ectopic papillae indicate the ray sublineage has been expressed. No other sensillum is known to generate a papilla similar to that observed; in *lin-22* hermaphrodites, no ectopic papillae were observed ($n=80$) in spite of the presence of ectopic postdeirid sublineages¹³. d, Similarity of representative sense organ sublineages in *C. elegans* and in *Drosophila*³⁰ suggests that these developmental programs were present in an early common ancestor of metazoans. In both organisms, peripheral sense organs develop from small clones of cells generated in a fixed division pattern from neuroblasts lying in the epidermis. Expression of the neuroblast sublineages is mediated by genes related to *achaete-scute* genes. N, neurons; sh, sheath cells; spt, support cells; X, programmed cell death. METHODS. Transgenic EM321 nematodes containing an EM#226 plus pRF4 mixed extrachromosomal array (Fig. 2, legend) were irradiated with 4,000 rads of γ -ray radiation. A strain (EM469) carrying a stably integrated transgene, *bxln1*, was isolated based on 100% transmission of the dominant Rol phenotype. EM469 hermaphrodites were mated to *him-5(e1490)* males and a Rol F2 hermaphrodite that segregated no *Lin-32* sons was isolated to establish a *him-5(e1490)*; *bxln1* line (EM470). For heat-shock experiments, populations of synchronized animals were obtained by feeding starvation-arrested L1 larvae isolated by hypochlorite treatment of gravid hermaphrodites. Heat shock was at 30 °C for 3 h. The time of heat shock giving the maximal number of ectopic ray papillae was determined to be 39–42 h after hatching, during the L4 larval stage.

series $+ > e1926 > bx46 > u282 > 0$, and that this causes a progressive decrease in the probability that a hypodermal cell will adopt the neuroblast cell fate.

To test whether the *lin-32* product is not only necessary but also sufficient to specify the neuroblast fate, we expressed *lin-32* cDNA ectopically under the control of a heat-shock promoter in both wild-type and *lin-32* mutant transgenic animals. In both cases we observed the generation of ectopic ray papillae in anterior body regions of both males and hermaphrodites, presumably due to the expression of ectopic ray sublineages (Fig. 3). Ray papillae are formed instead of rays whenever the ray sublineage is expressed outside the tail region of males (for example, in a *lin-22* mutant background¹³), presumably because of a lack of independent morphogenetic events necessary for ray extension¹⁴. Presence of ectopic papillae in anterior body regions of both males and hermaphrodites after heat shock suggests that if anterior seam cells in either sex are provided with LIN-32 from a transgene, they are capable of generating a neuronal sublineage and neurons. Evidently, these cells fail to generate neurons in wild type because of insufficient *lin-32* gene activity.

lin-32 is necessary for expression of the neuroblast cell fate by three different types of neuroblasts (Fig. 1b–d). In its absence, these cells become three different types of epidermal cells. The variable backgrounds in which *lin-32* acts, and the different consequences of its activity, must be determined by other genes.

Heterochronic genes may play a role in determining the type of neuroblast formed. The postdeirid neuroblast is transformed into a ray neuroblast in both males and hermaphrodites, when the L2 larval stage is transformed into the apparent equivalent of the L3 larval stage by a mutation in the heterochronic gene *lin-28* (ref. 15). Normally, the combined action of the genes *mab-5* and *lin-22* results in expression of the ray sublineage only in the tail¹⁶; these genes may be responsible for spatial regulation of *lin-32* activity. Appearance of ectopic papillae in hermaphrodites after heat shock indicates that regulation of the ray sublineage by sex-determination genes can be overcome by ectopic expression of *lin-32*, suggesting that *lin-32* may be a target of regulation by the sex-determination pathway. □

Received 3 October; accepted 11 November 1994.

- Weintraub, H. et al. *Science* **251**, 761–766 (1991).
- Lassar, A. B. et al. *Cell* **58**, 823–831 (1989).
- Skeath, J. B., Panganiban, G., Selegue, J. & Carroll, S. B. *Genes Dev.* **6**, 2606–2619 (1992).
- Chalfie, M. & Au, M. *Science* **243**, 1027–1033 (1989).
- Murre, C., Schouleber McCaw, P. & Baltimore, D. *Cell* **56**, 777–783 (1989).
- Jarman, A. P., Grau, Y., Jan, L. Y. & Jan, Y. N. *Cell* **73**, 1307–1321 (1993).
- Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y. & Jan, Y. N. *Nature* **369**, 398–400 (1994).
- Turner, D. L. & Weintraub, H. *Genes Dev.* **8**, 1434–1447 (1994).
- Tapscott, S. J. et al. *Science* **242**, 405–411 (1988).
- Hinz, U., Giebel, B. & Campos-Ortega, J. A. *Cell* **76**, 77–87 (1994).
- Hodgkin, J. *Nature* **304**, 267–268 (1983).
- Plenefisch, J. D., DeLong, L. & Meyer, B. J. *Genetics* **121**, 57–76 (1989).

13. Horvitz, H. R., Sternberg, P. W., Greenwald, I. S., Fixsen, W. & Ellis, H. M. *Cold Spring Harbor Symp. quant. Biol.* **43**, 453–463 (1983).
14. Sulston, J. E., Albertson, D. G. & Thomson, J. N. *Dev. Biol.* **78**, 542–576 (1980).
15. Ambros, V. & Horvitz, R. H. *Science* **226**, 409–416 (1984).
16. Kenyon, C. *Cell* **46**, 477–487 (1986).
17. Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110–156 (1977).
18. Baird, S. E., Fitch, D. H. A., Kassem, I. A. & Emmons, S. W. *Development* **113**, 515–526 (1991).
19. Gonzalez, F., Romani, S., Cubas, P., Modolelli, J. & Campuzano, S. *EMBO J.* **8**, 3553–3562 (1989).
20. Begley, C. G. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10128–10132 (1989).
21. Johnson, J. E., Birren, S. J. & Anderson, D. J. *Nature* **346**, 858–861 (1990).
22. Villares, R. & Cabrera, C. V. *Cell* **50**, 415–424 (1987).
23. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51**, 987–1000 (1987).
24. Krause, M., Fire, A., White Harrison, S., Priess, J. & Weintraub, H. *Cell* **63**, 907–919 (1990).
25. Kohl, N. E. et al. *Nature* **319**, 73–77 (1986).
26. Ruvkun, G. et al. *Genetics* **121**, 501–516 (1989).
27. Albertson, D. G. *Genetics* **134**, 211–219 (1993).
28. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. *EMBO J.* **10**, 3959–3970 (1991).
29. Palazzolo, M. J. et al. *Gene* **88**, 25–36 (1990).
30. Bodmer, R., Carretto, R. & Jan, Y. N. *Neuron* **3**, 21–32 (1989).

ACKNOWLEDGEMENTS. We thank D. Nguyen for isolating the *bx46* allele; A. Coulson for the YAC grids and cosmid; N. Baker and S. Baird for comments on the manuscript and members of our laboratory for discussion. Some nematode strains were received from the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. This work was supported by the US National Institutes of Health.

Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins

Wolfgang Seufert^{*†‡}, Bruce Futcher[†]
& Stefan Jentsch^{*†§}

^{*} Friedrich-Miescher-Laboratorium der Max-Planck-Gesellschaft, Spemannstrasse 37–39, 72076 Tübingen, Germany
[†] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

CELL cycle progression in eukaryotes is controlled by the *p34^{cdc2}/CDC28* protein kinase and its short-lived, phase-specific regulatory subunits called cyclins^{1,2}. In *Xenopus* oocytes, degradation of M-phase (B-type) cyclins is required for exit from mitosis and is mediated by the ubiquitin-dependent proteolytic system³. Here we show that B-type-cyclin degradation in yeast involves an essential nuclear ubiquitin-conjugating enzyme, UBC9. Repression of UBC9 synthesis prevents cell cycle progression at the G2 or early M phase, causing the accumulation of large budded cells with a single nucleus, a short spindle and replicated DNA. In *ubc9* mutants both CLB5, an S-phase cyclin^{4,5}, and CLB2, an M-phase cyclin^{6,7}, are stabilized. In wild-type cells the CLB5 protein is unstable throughout the cell cycle, whereas CLB2 turnover occurs only at a specific cell-cycle stage⁸. Thus distinct degradation signals or regulated interaction with the ubiquitin-protein ligase system may determine the cell-cycle specificity of cyclin proteolysis.

On the basis of its sequence similarity to known ubiquitin-conjugating (UBC) enzymes catalysing the covalent attachment of ubiquitin to proteolytic substrates⁹, we identified the *UBC9* gene encoding a novel member of this enzyme family from *Saccharomyces cerevisiae*. Interrupted by a single intron, the *UBC9* open reading frame encodes a protein (*M*, 18K, 157 amino acids; Fig. 1a) that shares ~35% identity with other UBC proteins⁹, including a conserved, putative active-site cysteine residue. UBC9 can form a thioester intermediate with ubiquitin *in vitro* (not shown) and substitution of the conserved cysteine by a serine residue resulted in a complete loss of UBC9 function *in vivo* (see later). We conclude that this cloned gene encodes a functional UBC enzyme.

a

```

-523 TCTAGAGCACTAATCAGTTTATTAATAATCTTCTGCTTCCATATTCTCGTTACCGTTA
-463 TTTTCATCAAAATTTGCGAACTCATTTTGCAAACTACCATCATTTCTTCAATCTGGG
-403 TCCTCTTCAGATTTCATAGTACCTGTACGCCATCACTGCTTCTTCTGCGCATCTCTC
-343 TCTTTGTTTTCATCATCTGTTGTACACAGAACCTTCGCTTCACATCATCGGCTCTGCT
-283 TCATCCAGTTTTCATGTTGTTCTTCTCTAAAAGGTCATCCAAATCATCAAAATATCGTAC
-223 TCGTTTTTCATTCATTACTCTGTTGTATGTTTGGCATTTCCTTTCCTGCAATACCTTC
-163 GGTTCGCCAATTTGTAATCTTCTTCTCACTTTATATCTCTCAGAAACCGGTTTAAACAT
-103 CTGGAATTAATAATTTCTCTGCTCCATAACAAACATTTAAAAAAGAGAGAAATTT

-43 AGCATAGGATAAGCACACACTGGCACCATTTTTGGGAAGCAATATGAGTAGTTTGTGTCT
MetSerSerLeuCysLe

18 ACACGCTCTTCAGGAAGAAAGGTAAGTAGTAGTTTCTCTCTTTTATGCTTACATTCTGT
uGlnArgLeuGlnGluGluAr

78 AGGCATACACAATTTTCATCCAGCGGTATACATAACAAATCGATGAACCTTAACCTGTTTAC

138 TTGAATAACAGAAAAAATGGAGAAAGGATCATCCATTGGATTTTATGCCAAACAGTT
glsLysTrpArgLysAspHisProPheGlyPheTyrAlaLysProVal

198 AAGAAAGCTGATGGGTCCATGGATTTCAGAAATGGGAAGCTGGTATCCAGGCAAGAA
LysLysAlaAspGlySerMetAspLeuGlnLysTrpGluAlaGlyIleProGlyLysGlu

258 GGTACAAACTGGGCGGTGGTGTGTACCCAAATACAGTCGAATATCCAAATGAATATCTCT
GlyThrAsnTrpAlaGlyGlyValTyrProIleThrValGluTyrProAsnGluTyrPro

318 TCAAAACCTCCAAAGGTTAAATTTCCAGCCGGATTTTATCATCCAAACGTGTATCCAAAT
SerLysProProLysValLysPheProAlaGlyPheTyrHisProAsnValTyrProSer

378 GGCACAATATGTTTAAATTTTAAATGAAGATCAAGATTGGAGACCGGCATCACGTTA
GlyThrIleCysLeuSerIleLeuAsnGluAspGlnAspTrpArgProAlaIleThrLeu

438 AAACAAATGTTTCTTGGGTTTCAGGATCTTTAGACTCTCCAAATCCAAATTCCTCTGCT
LysGlnIleValLeuGlyValGlnAspLeuLeuAspSerProAsnProAsnSerProAla

498 CAAGAGCCTGCATGGAGATCATTTTCAAGAAATAGGCGGAATATGACAAGAAAGTTTGT
GlnGluProAlaTrpArgSerPheSerArgAsnLysAlaGluTyrAspLysLysValLeu

558 CTTCAAGCTAAACAGTACTCTAAATAGAGGGAATCCATCTTTCCCATCTTCTCTCTTTT
LeuGlnAlaLysGlnTyrSerLys

618 GTACTTTTATTAATAATGTCGTTGTGTAACAAAAATAGAGCAAAATACATTATTTACA
AATTCTCAAAATAATTTTGTCTTCTTGTCTTATGCTAAGTAAATAGAAAGATATTT
678 TTTGTACCAATTTCTATAAGTATGGCACTATATACACTTTAATAATCTATTGGTTAGT
738 AGAATTTTCATTCATTTTGTAGTGAATGAACTAGCAACGTAGTAAGCAATCATGGCA
798 TCTTTCTTTTAGTTCGGGATTTTGTTTTATCAACATTTTGAATGTGCTCCTCAATTT
858

```

b

anti-βgal DAPI

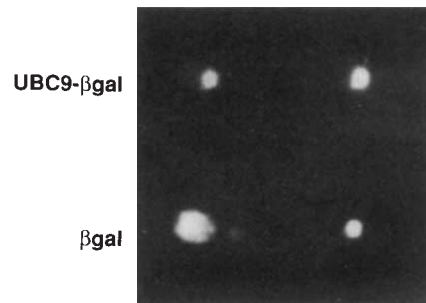


FIG. 1 Primary structure and intracellular localization of UBC9. **a**, Nucleotide and predicted amino-acid sequence of the *UBC9* gene (EMBL accession no. X82538). The position of the intron was found by comparing genomic and cDNA sequences. Splice sites and putative branch point are underlined. Nucleotide numbers starting at the first nucleotide of the coding region are given on the left. The cysteine residue required for UBC9-ubiquitin thioester formation is shown in bold; the proline residue changed to serine in *ubc9-1* is in italics. **b**, Nuclear localization of a UBC9 fusion protein. Cells expressing a UBC9-β-galactosidase fusion protein (UBC9-βgal) or unfused β-galactosidase (βgal) were immunostained with antibodies against βgal (anti-βgal). Staining with DAPI (4',6'-diamidine-2-phenylindole) shows the position of the nucleus.

METHODS. Standard genetic and molecular biology techniques were used²⁹. *UBC9* was cloned from a yeast genomic λEMBL3A library with a PCR-generated fragment amplified with degenerate oligonucleotide primers CGGAATTCGTTTAC/TGAAGGGGIGGTTT and GCTCTAGAATIG-TAA/GIGCIGGIG/CT/ACCA (I, inosine) corresponding to amino-acid sequences VYEGGVF and WSPALTI (single-letter code), conserved in many UBC proteins. *UBC9* cDNA was cloned by PCR with gene-specific primers from reverse-transcribed RNA. UBC9 was fused to βgal by inserting a 1.1-kb *Xba*-*Sal* fragment into YEp357R (refs available on request). Cells expressing this fusion or βgal from plasmid pUB23 (ref. 23) were fixed and immunostained with a mouse monoclonal antibody against βgal (Promega).

[†] Present addresses: Institut für Genetik, Universität München, Maria-Ward-Strasse 1a, 80638 München, Germany (W.S.); ZMBH, Universität Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany (S.J.).

[§] To whom correspondence should be addressed.

UBC enzymes function primarily in pathways of proteolysis^{9,11} and mediate different cellular functions^{12,18}. Whereas most known UBC genes are dispensable for life, *ubc9* mutants generated by gene disruption are inviable. Spores of *ubc9* disruptants are able to germinate and the cells die after a few divisions. We constructed a *ubc9* deletion strain kept alive by *UBC9* expression from the *GAL1* promoter. In the presence of glucose, *UBC9* transcription is turned off and cells stop dividing. The majority of cells arrest as large budded cells (Fig. 2a) with an increased cell volume and a single nucleus close to the neck. Flow cytometric analysis (FACS) showed that these cells had replicated their DNA. The presence of short, pre-anaphase spindles in *UBC9*-depleted cells (Fig. 2b) suggests that although these cells fail to go through mitosis, they execute early steps of spindle formation.

We generated a *ubc9* temperature-sensitive (ts) mutant (*ubc9-1*) by introducing a mutation equivalent to that of another mutant ts-ubiquitin-conjugating enzyme¹⁹. When shifted to 37 °C, *ubc9-1* mutants arrest with phenotypes similar to *UBC9*-depleted cells (not shown). Using this allele, we observed that the arrest phenotype of *ubc9-1* cells was not affected by the absence of *RAD9* (ref. 20), a gene that prevents cell division of cells containing incompletely replicated DNA. Furthermore, on hydroxyurea-containing plates, *ubc9-1* cells released from the *ubc9* cell-cycle block were able to go through another division (data not shown), indicating that the *ubc9* arrest occurs at a stage subsequent to the hydroxyurea execution point²¹. In the absence of noticeable metaphase markers in yeast, we conclude that *UBC9* is required after S phase at a stage defined as G2/M (ref. 6).

In *Xenopus* oocytes, exit from mitosis requires the ubiquitin-dependent degradation of mitotic (B-type) cyclins³. A *UBC9*- β -galactosidase fusion protein (which complements *ubc9* mutations) localizes to the nucleus (Fig. 1b) and consequently *UBC9* may act on nuclear proteins such as cyclins. We tested whether *UBC9* mediates the turnover of the (B-type) cyclins CLB5 (refs 4, 5) and CLB2 (refs 6, 7), which are maximally expressed at the onset of S phase and mitosis, respectively. To measure their stability during the cell cycle, CLB5 and CLB2 proteins harbouring an epitope tag were expressed from the *GAL* promoter in

synchronized cells and protein levels were followed after promoter shut-off. The CLB5 protein was essentially short-lived throughout the cell cycle, with a moderate variation of its half-life ranging from an estimated 5–10 min in α -factor-arrested G1 cells and 15–20 min in cells arrested in either S or M phase (Fig. 3a). In *ubc9-1* mutants, however, CLB5 was significantly stabilized and its half-life increased to ~1 h in M phase (Fig. 3b). *UBC9* can apparently mediate CLB5 turnover throughout the cell cycle as stabilization was also observed in non-synchronized cells (not shown). Confirmation was obtained in direct pulse-chase turnover measurements of a CLB5- β -galactosidase fusion followed by immunoprecipitation with β -galactosidase-specific antibodies. This fusion protein retains regular CLB5 activity *in vivo* and its short half-life was again found to be *UBC9*-dependent (Fig. 3c). Stabilization of CLB5 may be detrimental for normal cell proliferation as continuous high expression of CLB5 to levels still tolerated by wild-type cells strongly interfered with proliferation of *ubc9-1* mutants (Fig. 3d). Highly expressed CLB5 had a similar toxic effect in a proteasome-mutant strain (*pre1-1*) which is known to be compromised in the degradation of ubiquitinated proteins^{22,23}. We conclude that CLB5 turnover is mediated by the ubiquitin/proteasome pathway and involves the *UBC9* ubiquitin-conjugating enzyme.

In contrast to CLB5, stability of the M-phase cyclin CLB2 is strikingly regulated during the cell cycle. CLB2 is essentially stable during S and M phase (half-life >1 h), but extremely short-lived in pre-Start G1 cells (half-life <5 min) (Fig. 4a and ref. 8). To follow the stability of the CLB2 protein when cells arrest in G1 phase, wild-type and *ubc9-1* mutant cells that constitutively express CLB2 were treated with α -factor (Fig. 4b). Whereas RNA levels remained high, CLB2 protein fell below detectable levels in wild-type G1 cells. This was not observed in *ubc9-1* mutants where CLB2 remained sufficiently stable to accumulate after α -factor arrest (Fig. 4c), indicating that in addition to CLB5 proteolysis, *UBC9* is also involved in CLB2 turnover in pre-Start G1 cells. Failure to degrade CLB2 is expected to cause an anaphase arrest²⁴ (that is, with a long spindle). But stabilization of CLB5 and other B-type cyclins expressed earlier in the cell cycle² may contribute to the pre-

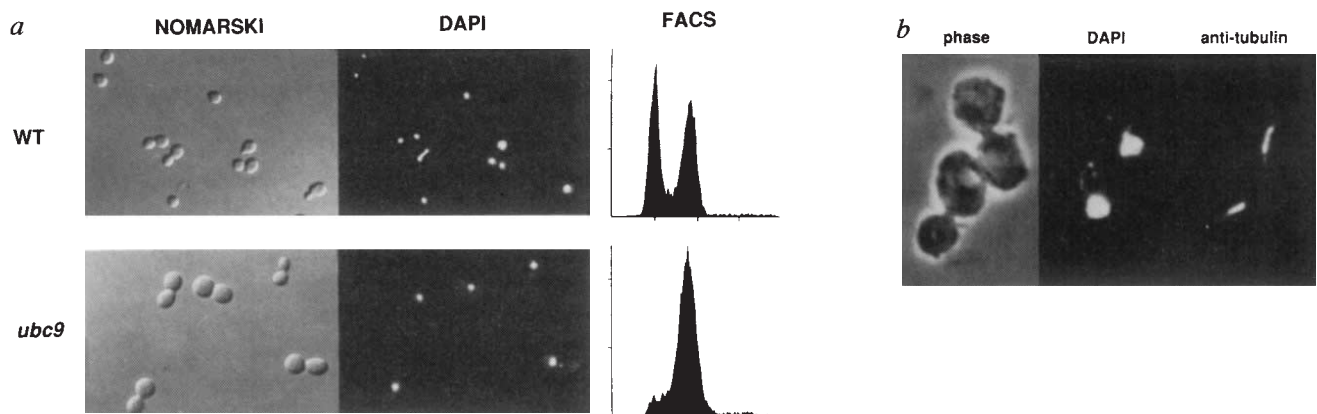


FIG. 2 Cell-cycle arrest phenotype conferred by *UBC9* depletion. *UBC9* synthesis was turned off in a strain carrying the gene under the control of the *GAL1* promoter. a, Micrographs showing cell morphology (Nomarski), stained nuclei (DAPI), and DNA content determined by flow cytometry (FACS) of wild type (WT) and cells depleted of *UBC9* (*ubc9*). The percentages of unbudded, small or large budded cells were, respectively, 48, 35 or 17% for wild type and 12, 6 or 82% for *UBC9*-depleted cells. The 2N DNA peak amounted to 46% (WT) or 84% (*ubc9*). b, Phase-contrast micrograph (phase) of *UBC9*-depleted cells is shown with the corresponding nuclei (stained with DAPI) and spindles immunostained with tubulin antibodies (anti-tubulin). The fractions of cells with either no visible, a short or long intranuclear spindle were, respectively, 66, 19 or 15% ($n=340$) for

isogenic wild type, and 16, 78 or 6% ($n=280$) for *UBC9*-depleted cells.

METHODS. Yeast strains used in this study are derivatives of DF5 (ref. 15). A DNA fragment with the *TRP1* marker inserted into *UBC9* (*NcoI* site at codon 36) was used for one-step gene disruption. Viability of this strain (YW077, *MAT α*) was rescued by a *GAL1* promoter-*UBC9* fusion gene (0.5 kb *Bam*HI cDNA fragment of *UBC9* in YlpG2) integrated at the *LEU2* locus. As a result of considerable overexpression of the stable *UBC9* protein in YPgal medium, cells went through about 8 divisions after repressing *UBC9* synthesis in YPD. Cells were collected, fixed in 70% ethanol and stained with DAPI for microscopy or processed for cytometry. A mouse monoclonal antibody against β -tubulin (Amersham) was used for *in situ* immunofluorescence.

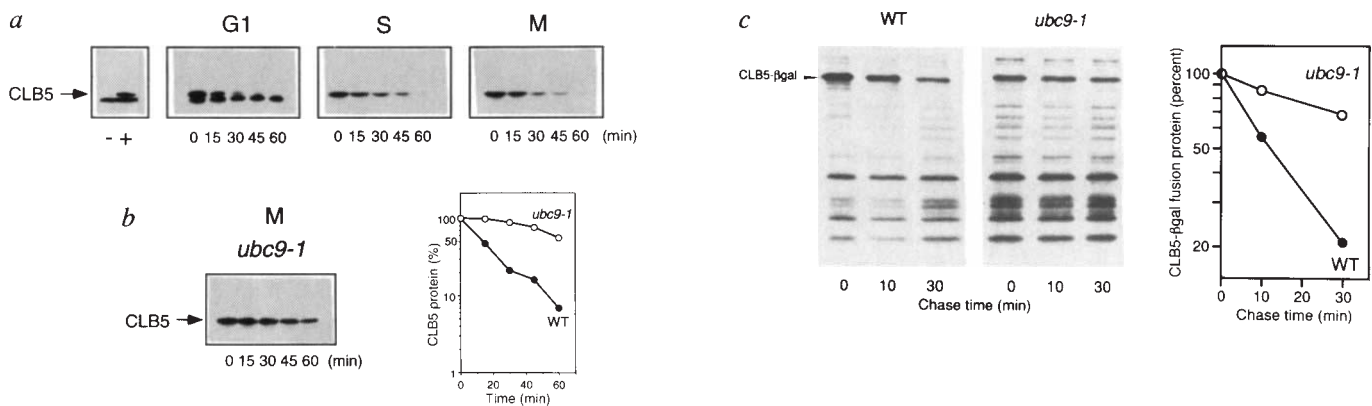


FIG. 3 CLB5 degradation involves UBC9. **a**, CLB5 stability during the cell cycle. Epitope-tagged CLB5 (left upper band in + lane; the lower crossreacting band is also present in cells bearing no (–) tagged CLB5) was expressed from the *GAL1* promoter in wild-type cells synchronized by α -factor (G1), hydroxyurea (S) or nocodazole (M); protein levels were followed on western blots after promoter shut-off. **b**, CLB5 stability in *ubc9-1* cells pre-arrested in M phase (M). Values obtained by scanning were plotted for half-life estimates. **c**, Stability of a CLB5- β -galactosidase (CLB5- β gal) fusion in wild type (WT) and *ubc9* mutant (*ubc9-1*) cells pre-arrested in M phase by nocodazole treatment. Protein turnover was analysed by pulse-chase immunoprecipitation²³. **d**, Overexpression of CLB5 is toxic to *ubc9-1* or proteasome mutants. Wild-type (*UBC9*, *PRE1*) and mutant strains (*ubc9-1*, *pre1-1* (ref. 23)) carrying a *GAL1* promoter-CLB5 fusion gene (pGAL1-CLB5) were spotted in serial dilutions on solid medium with carbon sources repressing (glucose) or inducing (galactose) CLB5 overexpression. Plates were incubated for 3 (glucose) or 5 (galactose) days at 28 °C.

METHODS. CLB5 on a 2.2-kb *HindIII* fragment was fused to the *GAL1* promoter in vector YCplac33. A triple HA1 epitope³⁰ was inserted between *BfrI* and *SpeI* sites at the CLB5 carboxy terminus. The CLB5- β gal fusion is a 2.1-kb *Asp718-NsiI* fragment of pGAL-CLB5 in YEp357. The temperature sensitive *ubc9-1* allele was constructed by introducing a mutation corresponding to the *cdc34-1* allele¹⁹. Codon 69 CCT (Pro) of *UBC9* was changed to TCT (Ser) using PCR-mediated *in vitro* mutagenesis. In strain YW0102 (*MATa*, *ubc9-Δ1::TRP1*, *LEU2::ubc9-1*) *UBC9*-coding sequences (*Clal-ScaI*) are replaced by the *TRP1* marker

and a 1.5-kb *XbaI-SspI* fragment expressing the *ubc9-1* allele is integrated at the *LEU2* locus. The wild-type strain was YW085 (*MATa*, *bar1::HIS3*). For cell-cycle arrest, wild-type cells were treated for 3 h at 30 °C with 50 ng ml⁻¹ α -factor, 0.1 M hydroxyurea or 15 μ g ml⁻¹ nocodazole; *ubc9-1* mutant cells were synchronized by nocodazole treatment at 25 °C before up-shift to 35 °C. Arrest was confirmed by microscopic inspection and FACS analysis. As high levels of CLB5 interfere with G1 arrest by α -factor, the regular induction (2% galactose for 12 h) was reduced to 90 min for the experiment shown in the G1 panel. The resulting low level of CLB5 accounts for the comparably strong crossreacting signal. Time courses were started by addition of glucose to 2% final concentration. Antibody 12CA5 and an ECL system (Amersham) were used for immunoblot analysis as described³⁰. Protein levels were quantified from suitable exposures with a laser densitometer.

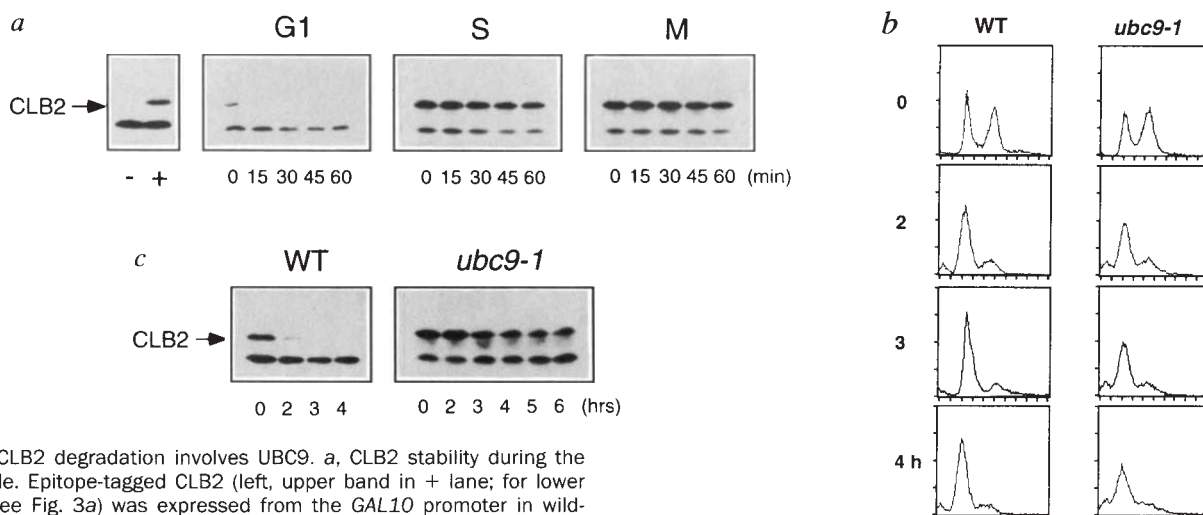


FIG. 4 CLB2 degradation involves UBC9. **a**, CLB2 stability during the cell cycle. Epitope-tagged CLB2 (left, upper band in + lane; for lower band, see Fig. 3a) was expressed from the *GAL10* promoter in wild-type cells arrested in G1, S and M phase (as in Fig. 3a) and protein levels were followed by western analysis after promoter shut-off. **b**, G1 cell-cycle arrest of wild type (WT) and *ubc9-1* cells after treatment with α -factor. FACS analysis shows the change of the DNA content of both strains after α -factor treatment, indicating that both strains arrest with similar kinetics. **c**, CLB2 levels in wild type and *ubc9-1* mutant cells during G1 arrest caused by α -factor treatment. Samples and time points are identical to those in **b**. RNA levels of tagged CLB2 expressed from a truncated CLB2 promoter were unaffected by the arrest.

METHODS. The triple HA1 epitope sequence³⁰ was added 3' to the CLB2 coding sequence. Tagged CLB2 on a 2.2-kb *EcoRI-BamHI* frag-

ment was fused to the *GAL10* promoter in vector YCplac33. CLB2 induction, cell-cycle arrest and turnover analysis were done as described in Fig. 3 legend. Levels of tagged CLB2 were followed in wild-type strain YW085 and *ubc9* mutant strain YW0103 (*MATa*, *bar1::HIS3*, *ubc9-Δ1::TRP1*, *LEU2::ubc9-1*) during treatment with α -factor at 30 °C. Constitutive CLB2 expression was conferred by a 620-bp CLB2 promoter fragment extending to the *XhoI* site. Differences in the relative intensities of CLB5 or CLB2 to the crossreacting protein reflect differences in their expression levels (Fig. 3 legend).

anaphase arrest (short spindle; Fig. 2b) of *ubc9* mutants. The relevant substrates may even include proteins other than known cyclins whose ubiquitin-dependent degradation is required for the metaphase-to-anaphase transition²⁵.

Cyclin-dependent kinases are thought to be the master regulators of the highly ordered events of the eukaryotic cell cycle. These events are accompanied by the ordered association of the catalytic subunit of the kinase with distinct phase-specific cyclins, suggesting that rapid dismantling of one kinase-cyclin complex may be required for another to be formed. Previous data³ indicated that this may be achieved by ubiquitin-dependent degradation of certain B-type cyclins. Our data indicate that both S- and M-phase B-type cyclins are destroyed by a common proteolysis pathway that involves UBC9. Further components involved in cyclin degradation are the proteasome (Fig. 3d) and certain ATPases²⁶ which may assist in proteasome function.

As CLB2 is highly stable in S or M phase when CLB5 is rapidly turned over, CLB2 proteolysis must be regulated at a level distinct from UBC9 action and subsequent steps in the ubiquitin pathway. Possible mechanisms are regulated stage-specific cyclin modification or recognition. Recognition by the destruction machinery may involve specific proteolytic signals within the cyclin molecule such as the 'destruction box' found in all B-type cyclins³ (including CLB5 and CLB2) and specific recognition proteins called ubiquitin ligases²⁷. Some proteins may bear several proteolytic signals that target the protein for degradation by distinct ubiquitination pathways²⁸. The residual instability of the cyclins (for example, CLB5; Fig. 3b, c) in *ubc9* mutant cells may also suggest that cyclins are degraded by more than one degradation pathway. Cell-cycle specificity of cyclin degradation could be conferred by distinct proteolytic signals and/or specific components of the ubiquitin-protein-ligase system.

Note added in proof: Recently we learned that the *S. pombe* gene *hus5*⁺ required for normal mitosis encodes a protein homologous to UBC9 (Khodairy *et al.*, *J. Cell Sci.*, in the press). □

Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade

Martin Scheffner*, Ulrike Nuber* & Jon M. Huibregtse†

* Deutsches Krebsforschungszentrum, Angewandte Tumoriologie, Im Neuenheimer Feld 242, 69120 Heidelberg, Germany

† Harvard Medical School, Department of Pathology, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

UBIQUITINATION of proteins involves the concerted action of the E1 ubiquitin-activating enzyme, E2 ubiquitin-conjugating enzymes and E3 ubiquitin-protein ligases¹⁻³. It has been proposed that E3s function as 'docking proteins', specifically binding substrate proteins and specific E2s, and that ubiquitin is then transferred directly from E2s to substrates¹⁻⁵. We show here that formation of a ubiquitin thioester on E6-AP, an E3 involved in the human papillomavirus E6-induced ubiquitination of p53 (refs 6-10), is an intermediate step in E6-AP-dependent ubiquitination. The order of ubiquitin transfer is from E1 to E2, from E2 to E6-AP, and finally from E6-AP to a substrate. This cascade of ubiquitin thioester complexes suggests that E3s have a defined enzymatic activity and do not function simply as docking proteins. The cysteine residue of E6-AP responsible for ubiquitin thioester formation was mapped to a region that is highly conserved among several proteins of unknown function, suggesting that these proteins share the ability to form thioesters with ubiquitin.

E6-AP-dependent ubiquitination requires a distinct E2 activity¹⁰, which is represented by members of a subfamily of E2s that includes *Arabidopsis thaliana* (At) UBC8 (ref. 11) and human UbcH5 (ref. 12). Coprecipitation analyses, however, revealed no complex between E6-AP and AtUBC8 or UbcH5 (data not shown). This may indicate that the interaction between E6-AP and these E2s is weak and cannot be detected under the conditions of a coprecipitation experiment. Because a necessary step in the ubiquitination of proteins is the E1-dependent attachment of ubiquitin to E2s via a thioester bond^{1,2,13}, we tested the effect of E6-AP on the formation of a ubiquitin thioester on AtUBC8 (Fig. 1). In the presence of E1, a product with the characteristics of AtUBC8 linked to ubiquitin by a thioester bond was detected. Complex formation of E2 with ubiquitin was not significantly affected by the presence of E6-AP, but in the presence of E6-AP an additional complex containing ubiquitin was observed migrating at an approximate relative molecular mass of 130,000 (130K; Fig. 1a). This complex, like the UBC8-ubiquitin complex, had the characteristics of a thioester adduct in that it was labile to reducing agents such as dithiothreitol (DTT) (Fig. 1b) and disappeared on treatment with 0.1 M NaOH but not with 1 M formic acid (data not shown). The ubiquitin used was a glutathione-S-transferase ubiquitin fusion protein (GST-ubiquitin) of M_r 34K¹⁰. The molecular mass of the recombinant form of E6-AP used in this experiment is 95K⁹, suggesting that the 130K complex might consist of a single molecule each of E6-AP and GST-ubiquitin. To test this possibility, an amino-terminally truncated form of E6-AP with an M_r of 76K, which functions in E6-AP-dependent ubiquitination⁸⁻¹⁰, was examined for ubiquitin adduct formation. The 76K form of E6-AP and GST-ubiquitin formed a complex of ~110K with the characteristics of a thioester (Fig. 1c). Furthermore, as shown previously for E6-AP-dependent protein ubiquitination¹⁰, E6-AP thioester formation was dependent on the presence of E1 and AtUBC8 or UbcH5 (data not shown). The only other proteins known to form thioesters with ubiquitin are E1 (refs 14, 15) and the members of the E2 family^{16,17}, although it was previously noted that E3s might also

Received 23 August; accepted 11 November 1994.

- Murray, A. & Kirschner, M. W. *Science* **246**, 614-621 (1989).
- Nasmyth, K. *Curr. Opin. Cell Biol.* **5**, 166-179 (1993).
- Glötzer, M., Murray, A. W. & Kirschner, M. W. *Nature* **349**, 132-138 (1991).
- Epstein, C. & Cross, F. *Genes Dev.* **6**, 1695-1706 (1992).
- Schwob, E. & Nasmyth, K. *Genes Dev.* **7**, 1160-1175 (1993).
- Surana, U. *et al. Cell* **65**, 145-161 (1991).
- Ghiara, J. B. *et al. Cell* **65**, 163-174 (1991).
- Amon, A., Irmiger, S. & Nasmyth, K. *Cell* **77**, 1037-1050 (1994).
- Jentsch, S. *A. Rev. Genet.* **26**, 177-205 (1992).
- Finley, D. & Chau, V. *A. Rev. Cell Biol.* **7**, 25-69 (1991).
- Hershko, A. & Ciechanover, A. *A. Rev. Biochem.* **61**, 761-807 (1992).
- Jentsch, S., McGrath, J. P. & Varshavsky, A. *Nature* **329**, 131-134 (1987).
- Goebel, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. & Byers, B. *Science* **241**, 1331-1335 (1988).
- Seufert, W. & Jentsch, S. *EMBO J.* **9**, 543-550 (1990).
- Seufert, W., McGrath, J. P. & Jentsch, S. *EMBO J.* **9**, 4535-4541 (1990).
- Wiebel, F. & Kunau, W.-H. *Nature* **359**, 73-76 (1992).
- Sommer, T. & Jentsch, S. *Nature* **365**, 176-179 (1993).
- Jungmann, J., Reins, H.-A., Schobert, C. & Jentsch, S. *Nature* **361**, 369-371 (1993).
- Ellison, K. S., Gwozd, T., Prendergast, J. A., Paterson, M. C. & Ellison, M. J. *J. Biol. Chem.* **266**, 24116-24120 (1991).
- Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317-322 (1988).
- Hartwell, L. H. *J. molec. Biol.* **104**, 803-817 (1976).
- Heinemeyer, W., Kleinschmidt, J. A., Saldowsky, J., Escher, C. & Wolf, D. H. *EMBO J.* **10**, 555-562 (1991).
- Seufert, W. & Jentsch, S. *EMBO J.* **11**, 3077-3080 (1992).
- Surana, U. *et al. EMBO J.* **12**, 1969-1978 (1993).
- Holloway, S. L., Glötzer, M., King, R. W. & Murray, A. W. *Cell* **73**, 1393-1402 (1993).
- Ghiislain, M., Udvardy, A. & Mann, C. *Nature* **366**, 358-362 (1993).
- Hershko, A. *et al. J. Biol. Chem.* **269**, 4940-4946 (1994).
- Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. *Cell* **74**, 357-369 (1993).
- Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology* (Green and Wiley, New York, 1994).
- Tyers, M., Tokiwa, G., Nash, R. & Fletcher, B. *EMBO J.* **11**, 1773-1784 (1992).

ACKNOWLEDGEMENTS. We thank G. Smith for FACS analysis of UBC9-depleted cells; F. Cross, K. Nasmyth, L. Prakash and S. Reed for strains and plasmids; U. Ehringer for technical assistance; A. Amon, K. Nasmyth and E. Schwob for information before publication; and S. Smith for comments on the manuscript. W. S. is a Heisenberg fellow of the Deutsche Forschungsgemeinschaft (DFG); B.F. and S.J. are supported by a grant from Human Frontier Science Program, and S.J. by grants from DFG and German-Israeli Foundation of Research and Development (GIF).

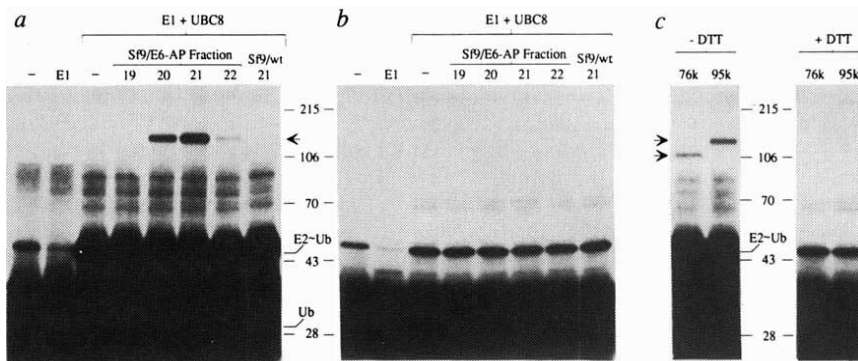
have such an activity¹⁸. The regions in E1 and the E2s that are responsible for thioester formation have been well characterized^{2,15}, but they show little similarity at the protein sequence level. Similarly, protein sequence comparisons did not reveal any significant similarity of E6-AP to either E1 or any of the E2s.

To determine whether thioester formation of E6-AP with ubiquitin correlated with the ability of E6-AP to function as an E3, a deletion mutant of E6-AP lacking the carboxy-terminal 84 amino acids, which can associate with E6 and p53 but cannot induce ubiquitination of p53 (ref. 9), was assayed for thioester formation (Fig. 2b). This truncated E6-AP could not form a thioester with ubiquitin (Fig. 2c), suggesting that the cysteine residue involved in thioester formation may lie in the C terminus of E6-AP. The 84 C-terminal amino acids contain only one cysteine residue, at position 833 (for numbering, see ref. 8). When this cysteine of the 95K form of E6-AP was altered to either serine or alanine, both mutant E6-APs could bind to p53 in the presence of E6 (data not shown), but could not induce ubiquitination of p53 (Fig. 2b) and did not form a thioester with ubiquitin (Fig. 2c). Thus, the cysteine residue at position 833 of E6-AP is essential for thioester formation as well as for E6-AP-dependent ubiquitination.

Our results indicate that the formation of a ubiquitin thioester on E6-AP is an intermediate step in E6-AP-dependent ubiquitination, and that the order of ubiquitin transfer is from E1 to E2 and then from E2 to E6-AP. This cascade of ubiquitin thioester complexes suggests that ubiquitin is transferred directly from E6-AP to the substrate. In support of this hypothesis, pulse-chase experiments revealed that radioactively labelled

FIG. 1 E6-AP forms a thioester with ubiquitin. *a*, Thioester formation of ubiquitin with baculovirus-expressed E6-AP. Thioester reactions contained ³²P-labelled GST-ubiquitin¹⁰, E1, *Arabidopsis thaliana* (At) UBC8, ATP and Mono-Q fractions (see below) containing different amounts of the 95K form of E6-AP. Fraction 21 contained the peak of E6-AP activity. As a control, the corresponding fraction of extracts from wild-type baculovirus-infected Sf9 cells was used. After 5 min at 25 °C, reactions were stopped in the absence of a reducing agent and subjected to SDS-PAGE followed by autoradiography. Because of the low amount of E1 used, the formation of a thioester between ubiquitin and E1 was not detectable. *b*, As *a*, but reactions were stopped in the presence of 100 mM DTT. *c*, Approximately equal amounts (150–200 ng) of the 95K or the 76K form of E6-AP⁹ were tested in a thioester assay. Reactions were either stopped in the absence of a reducing agent (–DTT) or in its presence (+DTT). The running positions of GST-ubiquitin (Ub) and the thioester adduct of E2 are indicated. The ubiquitin thioester adducts of the different forms of E6-AP are marked by arrows. Note that formation of a ubiquitin thioester on E6-AP is also observed with native ubiquitin (data not shown).

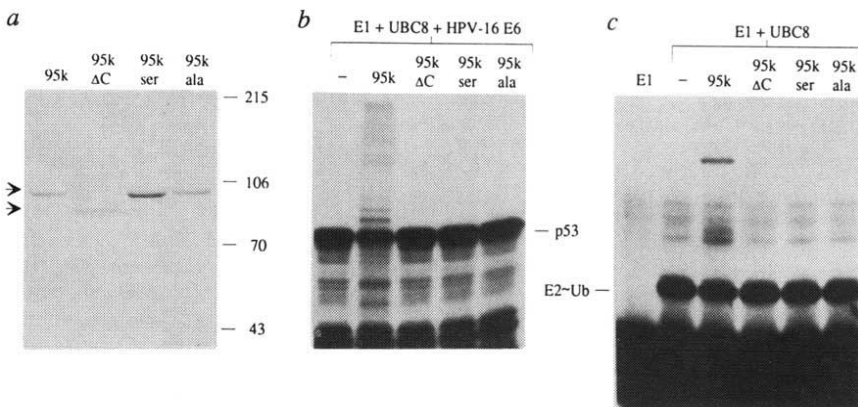
METHODS. Protein extracts were prepared 48 h after infection and fractionated by chromatography on Mono-Q¹⁰. As Sf9 cells contain no



detectable E6-AP activity⁹, fractions containing E6-AP were determined by the ability of these fractions to reconstitute ubiquitination of an E6-E7 fusion protein¹⁰. Thioester reactions contained approximately 1 μg ³²P-labelled GST-ubiquitin, 5 ng of E1 (ref. 15) and 200 ng of AtUBC8 (ref. 11) in 20 mM Tris-HCl pH 7.6, 50 mM NaCl, 4 mM ATP, 10 mM MgCl₂, 0.2 mM DTT. Reactions were terminated either by incubating the mixtures for 15 min at 30 °C in 50 mM Tris-HCl pH 6.8, 2% SDS, 4 M urea, 10% glycerol or by boiling the mixtures in the above buffer containing 100 mM DTT instead of urea. The whole reaction mixtures were separated on 10% SDS-polyacrylamide gels at 4 °C.

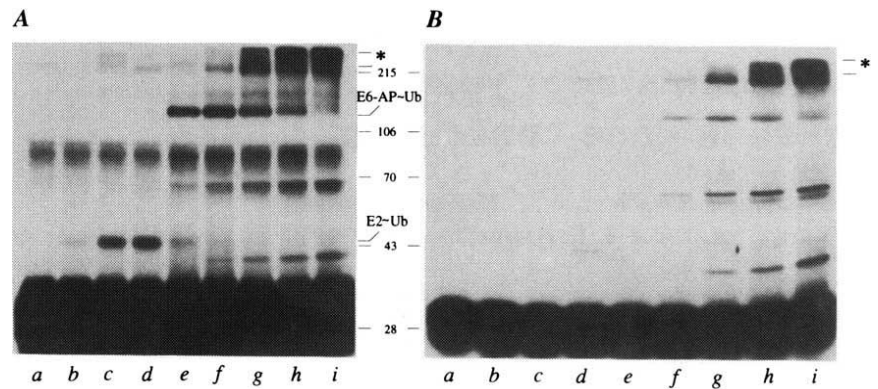
FIG. 2 A cysteine residue in the C terminus of E6-AP is essential for thioester formation. *a*, Coomassie stain of different mutant forms of E6-AP. Mono-Q fractions containing the 95K form of E6-AP, a truncated form of the 95K E6-AP lacking the C-terminal 84 amino acids (95kΔC), or mutants of the 95K E6-AP in which the cysteine at position 833 was altered to either serine (95k-ser) or alanine (95k-ala) were separated by SDS-PAGE and proteins stained with Coomassie blue. The running positions of *M_r* markers are indicated. The different forms of E6-AP are marked by arrows. *b*, E6-AP-dependent ubiquitination of p53. Approximately equal amounts of the different forms of E6-AP were tested for their ability to mediate the ubiquitination of p53 in the presence of the HPV-16 E6 protein¹⁰. *c*, Thioester formation of the mutated forms of E6-AP.

Approximately equal amounts of the different E6-AP were tested in the thioester assay as described for Fig. 1. Reactions were stopped in the absence of a reducing agent. For some E2s, it has been shown that replacement of the active-site cysteine by serine results in the formation of an ester with ubiquitin²⁰ which is insensitive to reducing agents. This property, however, is not observed with all members of the E2 family (S. Jentsch, personal communication). The serine



mutant of E6-AP did not form a detectable ester adduct with ubiquitin. METHODS. The construction of the recombinant baculovirus expressing the C-terminally truncated form of the 95K E6-AP has been described⁹. The cysteine residue at position 833 of the E6-AP open reading frame⁸ was altered to either serine or alanine by polymerase chain reaction directed mutagenesis. Recombinant baculoviruses expressing these mutated forms of E6-AP were generated as described⁹.

FIG. 3 Ubiquitin can be directly transferred from E6-AP to substrate proteins. It was previously reported that the Mono-Q fraction containing E6-AP also contains proteins that can serve as substrates in E6-AP-dependent ubiquitination¹⁰. The ubiquitination of these proteins was studied in a pulse-chase experiment. **A**, Reactions were terminated in the absence of a reducing agent. **B**, Reactions were terminated in the presence of 100 mM DTT. E1, AtUBC8, E6-AP and ³²P-labelled GST-ubiquitin were incubated for 3 min at 25 °C, resulting in the formation of ubiquitin thioesters on E6-AP (pulse) (lane e). Because of the low amounts of E1 and AtUBC8 used, thioester complexes of AtUBC8 with ubiquitin were barely detectable in the presence of E6-AP. After 3 min a 50-fold excess of unlabelled native ubiquitin was added and the reactions were further incubated for different times (chase) (lane e, 0 min; lane f, 1 min; lane g, 3 min; lane h, 5 min; lane i, 10 min). Addition of unlabelled native ubiquitin resulted in a significant decrease in the amount of detectable E6-AP thioesters, with a concomitant appearance of stably ubiquitinated proteins (marked by a star) migrating at the top of the gel. This strongly suggests that ubiquitin can be transferred directly from E6-AP to substrate proteins, although it is not clear whether the final transfer is mediated by E6-AP alone or by a complex consisting of E6-AP and AtUBC8. Running positions of *M_r* markers, of the thioester of E2 with ubiquitin, and of the thioester of E6-AP with ubiquitin are



indicated. The identity of the bands migrating at approximately 40, 66 and 135K is unknown (see also ref. 10), but the appearance of these bands is dependent on E6-AP. Lane a, after preincubation of E1 with unlabelled ubiquitin for 3 min, ³²P-labelled GST-ubiquitin and AtUBC8 were added and the reaction was incubated for an additional 10 min; lane b, same as lane a except that E1 was preincubated with ³²P-labelled GST-ubiquitin and then unlabelled ubiquitin and AtUBC8 were added; lane c, E1 and AtUBC8 were incubated with ³²P-labelled GST-ubiquitin for 3 min; lane d, same as lane c but after 3 min, unlabelled ubiquitin was added and the reaction incubated for an additional 10 min.

<i>hs e6ap</i>	PDTERLPTSTHCFNVLLPEYSSKEKLKERLLKAITYAK-GFGML
<i>hs 220kDa</i>	NPDDFLPSVMTVCYNLYKLDPYSSIEIMREKLLIAAREGQSQSFHLS
<i>hs 115kDa</i>	SGEYLPVAHTCYNLLDLPKYSSKEILSARLTQALDNYE-GFSLA
<i>hs 120kDa</i>	SDLERLPTASTCMNLLKLPEFYDETLLRSKLLYAIECAA-GFELS
<i>rat100kDa</i>	PDDQHLPTANTCISRLYVPLYSSKQLKQLLLAIKTKNFGFV
<i>sc 168kDa</i>	TADBYLPSVMTCANLYKLPKYTSKDIMRSRLCQAIEGAGAFLLS
<i>sc 90kDa</i>	GEVQQLPKSHTCFNVRDLPQYVDYDSMKQKLTLAVEETI-GFGQE

FIG. 4 The region of E6-AP involved in ubiquitin thioester formation is conserved among several proteins from different organisms. A 45-amino-acid overlap that encompasses the C termini of these proteins is shown. Amino acids that are conserved among at least five of the seven proteins are shown in bold type. The cysteine residue of E6-AP involved in ubiquitin thioester formation is marked with an asterisk. The sequences of the rat 100K and *S. cerevisiae* 168K have been published^{21,22}, while the others were retrieved from the EMBL data library (accession numbers D28476, D25215 and D13635 for *Homo sapiens* 220K, 115K and 120K, respectively and S53418 for *S. cerevisiae* 90K).

ubiquitin can be chased from E6-AP thioesters to substrate proteins on addition of excess amounts of unlabelled ubiquitin (Fig. 3). Therefore, it seems that E3s do indeed have ubiquitin-protein ligase activity and do not function simply as adaptors between substrate proteins and E2s as previously assumed. Further studies, however, should evaluate whether the proposed mechanism for E6-AP-dependent ubiquitination applies to E3-dependent ubiquitination in general.

Only two E3 genes have been isolated so far, those encoding UBR1 of yeast and E6-AP^{8,19}. These proteins have no apparent

sequence similarity but databank searches have shown that the C-terminal regions of several proteins from different eukaryotic organisms share significant homology to the C terminus of E6-AP (Fig. 4). The position of the cysteine residue that is essential for thioester formation and for E6-AP-dependent protein ubiquitination, as well as several surrounding residues, are conserved among all these proteins. Although the function of these proteins is unknown, it is tempting to speculate that they share the ability to form thioesters with ubiquitin and are members of an E6-AP like family of ubiquitin-protein ligases. □

Received 21 October; accepted 21 November 1994.

- Hershko, A. & Ciechanover, A. A. *Rev. Biochem.* **61**, 761–807 (1992).
- Jentsch, S. A. *Rev. Genet.* **26**, 177–205 (1992).
- Varshavsky, A. *Cell* **69**, 725–735 (1992).
- Dohmen, R. J., Madura, K., Bartel, B. & Varshavsky, A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7351–7355 (1991).
- Watkins, J. F., Sung, P., Prakash, S. & Prakash, L. *Genes Dev.* **7**, 250–261 (1993).
- Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. *Cell* **63**, 1129–1136 (1990).
- Huibregtse, J. M., Scheffner, M. & Howley, P. M. *EMBO J.* **13**, 4129–4135 (1991).
- Huibregtse, J. M., Scheffner, M. & Howley, P. M. *Molec. cell. Biol.* **13**, 775–784 (1993).
- Huibregtse, J. M., Scheffner, M. & Howley, P. M. *Molec. cell. Biol.* **13**, 4918–4927 (1993).
- Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. *Cell* **75**, 495–505 (1993).
- Girod, P.-A., Carpenter, T. P., van Nocker, S., Sullivan, M. L. & Vierstra, R. D. *Pl. J.* **3**, 545–552 (1993).
- Scheffner, M., Huibregtse, J. M. & Howley, P. M. *Proc. natn. Acad. Sci. U.S.A.* **91**, 8797–8801 (1994).

- Haas, A. L., Warms, J. V. B., Hershko, A. & Rose, I. A. *J. biol. Chem.* **257**, 2543–2548 (1982).
- Ciechanover, A., Elias, S., Heller, H. & Hershko, A. *J. biol. Chem.* **257**, 2537–2542 (1982).
- Hatfield, P. M. & Vierstra, R. D. *J. biol. Chem.* **267**, 14799–14803 (1992).
- Pickart, C. M. & Rose, I. A. *J. biol. Chem.* **260**, 1573–1581 (1985).
- Haas, A. L. & Bright, P. M. *J. biol. Chem.* **263**, 13258–13267 (1988).
- Berleth, E. S. *et al. J. biol. Chem.* **267**, 16403–16411 (1992).
- Bartel, B., Wüning, I. & Varshavsky, A. *EMBO J.* **9**, 3179–3189 (1990).
- Sung, P., Prakash, S. & Prakash, L. *J. molec. Biol.* **221**, 745–749 (1991).
- Müller, D., Rehbein, M., Baumeister, H. & Richter, D. *Nucleic Acids Res.* **20**, 1471–1475 (1992).
- Pascolo, S. *et al. Yeast* **8**, 987–995 (1992).

ACKNOWLEDGEMENTS. We thank P. M. Howley, S. Jentsch and K. Münger for critical reading of the manuscript. M.S. thanks S. Stephan for technical assistance and E.-M. DeVilliers and H. zur Hausen for their support. M.S. and J.M.H. thank P. M. Howley for his support. This work was supported by the Deutsche Forschungsgemeinschaft.

DNA structure-dependent requirements for yeast *RAD* genes in gene conversion

N. Sugawara, E. L. Ivanov, J. Fishman-Lobell*,
B. L. Ray*, X. Wu & J. E. Haber†

Rosenstiel Basic Medical Sciences Research Center and
Department of Biology, Brandeis University, Waltham,
Massachusetts 02254-9110, USA

In *Saccharomyces cerevisiae*, HO endonuclease-induced mating-type (*MAT*) switching is a specialized mitotic recombination event in which *MAT* sequences are replaced by those copied from a distant, unexpressed donor (*HML* or *HMR*)^{1,2}. The donors have a chromatin structure inaccessible for both transcription and HO cleavage^{1,2}. Here we use physical monitoring of DNA to show that *MAT* switching is completely blocked at an early step in recombination in strains deleted for the DNA repair genes *RAD51*, *RAD52*, *RAD54*, *RAD55* or *RAD57*. We find, however, that only *RAD52* is required when the donor sequence is simultaneously not silenced and located on a plasmid. *RAD51*, *RAD54*, *RAD55* and *RAD57* are still required when the same transcribed donor is on the chromosome. We conclude that recombination *in vivo* occurs between DNA molecules in chromatin, whose structure significantly influences the outcome. *RAD51*, *RAD54*, *RAD55* and *RAD57* are all required to facilitate strand invasion into otherwise inaccessible donor sequences.

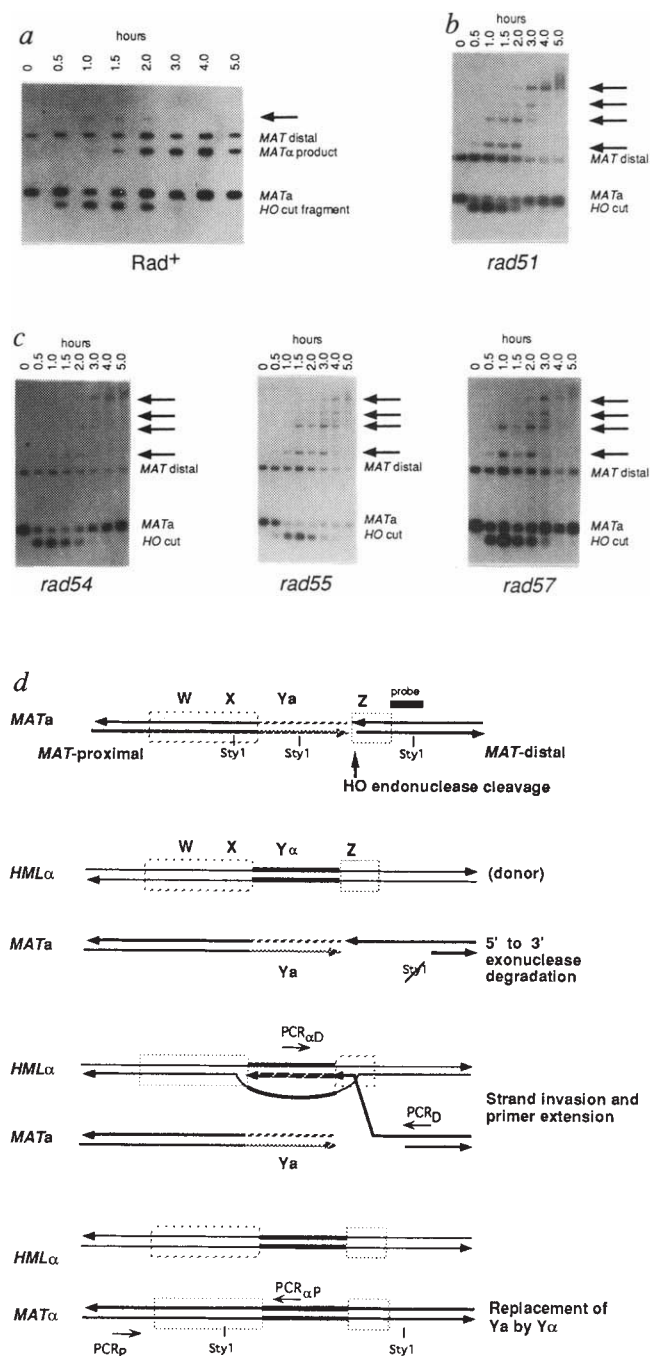
By physical monitoring of double-strand break (DSB)-induced recombination, the fate of recombining DNA can be observed in real time. A DSB induced at *MAT-Z* (Fig. 1) is repaired by copying intact mating-type sequences from an unexpressed donor locus, *HMLα* or *HMRα*. *MAT* switching takes one hour to complete, during which time several intermediate steps can be identified (Fig. 1d)³⁻⁶.

Null mutations in any of the recombination-repair genes *RAD51*, *RAD52*, *RAD54*, *RAD55* and *RAD57* are lethal in cells attempting to switch *MAT*^{4,7-9} (Fig. 1), but nothing is known about which specific events are affected. Switching from *MATα* to *MATα* in *Rad*⁺ strain R166 is shown in Fig. 1a. *MATα* product appears 1 h after *MATα* is cut. At intermediate times, higher-molecular-weight bands are seen that result from 5' to 3' degradation of the DSB⁴. In contrast, a *rad51* mutation prevents the appearance of product but instead exhibits extensive formation of single-stranded DNA where the DSB end is resected by as much as 20 kilobases (kb) (Fig. 1b). Indistinguishable results were obtained for *rad54*, *rad55* and *rad57* (Fig. 1c) and *rad52* (ref. 4).

We used a polymerase chain reaction (PCR)-based assay to determine whether these mutations prevented early steps of recombination. Strand invasion of 3'-ended *MAT-Z* DNA into *HMLα* and the initiation of copying donor information can be detected by PCR amplification^{4,5} (primers *PCR_{αD}* and *PCR_D*; Fig. 1d). This PCR product was not seen using the DNA samples from any of the *rad* mutants, although it was detected as early as 30 min after HO induction in a *Rad*⁺ strain (data not shown). Thus all of these mutations block *MAT* switching at an early step.

One puzzling feature of *MAT* switching is that ends of HO-cleaved *MAT* DNA are able to invade the same region of *HML* or *HMR* where HO cannot cut. Some of the *RAD* genes might be required to facilitate strand invasion into otherwise inaccessible regions of DNA. We moved *MAT* and the silent *HMR* locus to

FIG. 2 Effects of donor silencing on *MAT* switching in a centromeric plasmid. a, pNSU178 contains *MATα* and the silent *HMRα* locus in inverted orientation on a centromere plasmid marked by *LEU2* and *URA3*. b, A DSB was created in *MATα* by inducing the HO endonuclease as described in Fig. 1. The DNA from strains tNS908 (*rad51*) and tNS915 (*Rad*⁺) was similarly extracted and digested with *StyI* and run on a neutral agarose gel. c, *StyI* cuts within *MATα* but not *Yα*, thus yielding a distinctive *MATα* product band at 0.9 kb when *MATα* switches to *MATα*. The probe also hybridizes to the 1.3 kb *MAT_{distal}* band from pNSU178. Although *MATX*, *Y* and *Z* regions have been deleted from chromosome III, the chromosome retains two distal *StyI* fragments that hybridize with the probe. One comigrates with the pNSU178 *MAT_{distal}* fragment. The other migrates at 2.9 kb. d, Plasmid pNSU216 is a centromeric plasmid carrying *MATα* and an expressed donor sequence, *MATα-inc*, in inverted orientation. *MATα-inc* cannot be cleaved by HO. e, The HO endonuclease was induced with galactose to cut at the *MATα* sequence on pNSU216. The DNA was digested with *HhaI* and separated



* Present addresses: Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA (J.F.-L.); SymBio Tech Inc., Woburn, Massachusetts 01801, USA (B.L.R.).

† To whom correspondence should be addressed.

on a neutral agarose gel. The Southern blot was probed with the *MAT_{distal}* sequence shown in panel F. In strain *tNS900*, the *HO* endonuclease cuts the *MAT* sequences located on pNSU216 and on chromosome III. In the *HhaI* digest shown, pNSU216 gives rise to a 1.3-kb band (*MAT_α*, plasmid) and a 3.3-kb *MAT_α-inc* band. *HO* endonuclease cleaves *MAT* immediately adjacent to the *HhaI* restriction site such that the *HO* cut fragment is not distinguishable from the *HhaI* restriction fragment. When the *HO*-cut *MAT_α* is gene converted to *MAT_α-inc*, it simultaneously loses the *HhaI* restriction site overlapping the *HO* cut site and becomes a 2.2-kb restriction fragment (*MAT_α-inc* product). The chromosomal *MAT_a* sequence (2.6 kb) is also converted to *MAT_α-inc*, yielding a 2.7-kb restriction fragment. The 30% viability of these strains is explained by the fact that the single available *MAT_α-inc* donor could apparently be used only once, either by the cut *MAT* on the plasmid or the cut *MAT* on the chromosome. Cells that had repaired the chromosomal break almost always had lost the plasmid.

METHODS. Galactose induction of the *HO* endonuclease, carried on plasmid pFH800 (ref. 17), was performed as described for Fig. 1. pNSU178 was derived from plasmid pHF1 which consists of a pBR322 vector sequence with *MAT_α* (*EcoRI*–*HindIII*), *LEU2* (*PstI* fragment) and *HMR_a* (*PstI*–*EcoRI*), as well as *URA3* inserted at the *HindIII* site and *CEN3* inserted at the *BamHI* site. pNSU178 was created by inverting the orientation of the *SalI* fragment of pHF1 so that *MAT_α* and *HMR_a* are in inverted orientation. Plasmid pNSU215 (not shown) contains the *EcoRI*–*HindIII* sequence from *MAT_α* inserted between the *EcoRI* and *HindIII* sites of the *URA3*-marked centromere plasmid YCp50. pNSU216 was constructed from pNSU215 by inserting the *EcoRI*–*HindIII* fragment (filled in) containing *MAT_α-inc* into the *SmaI* site of *URA3*. The *rad* deletions are listed in Fig. 1 legend.

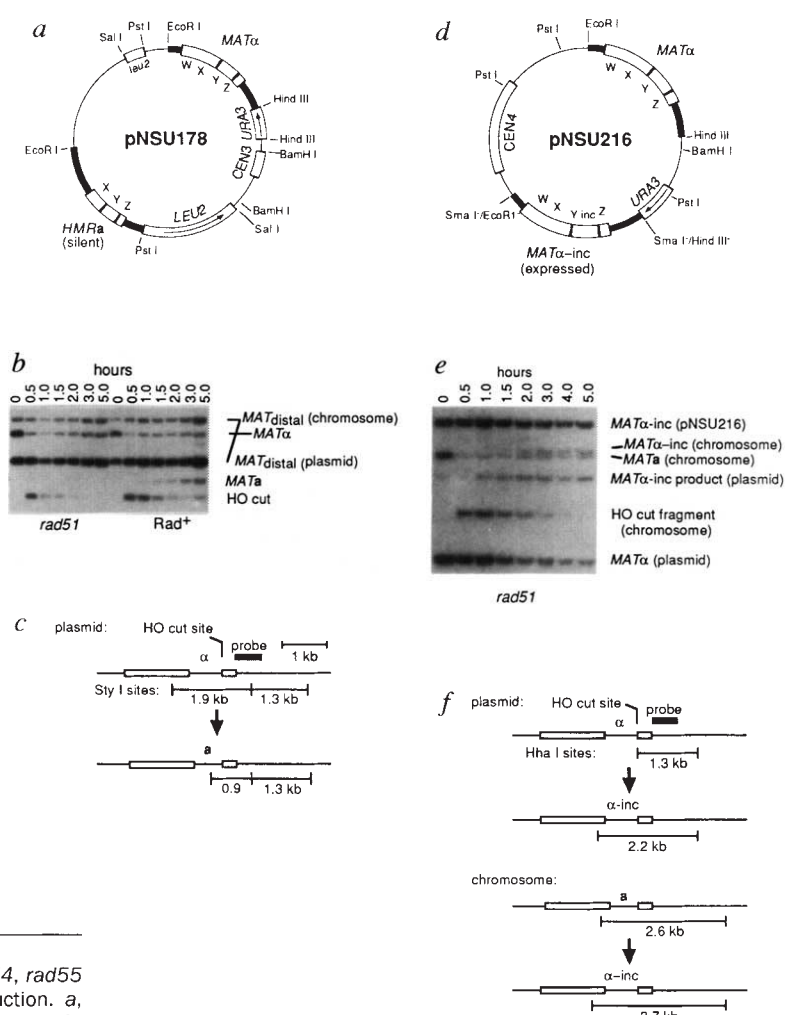


FIG. 1 Analysis of mating type switching in *Rad⁺*, *rad51*, *rad54*, *rad55* and *rad57* strains by Southern hybridization after *HO* induction. **a**, Strains R166 (*Rad⁺*); **b**, RY583T1 (*rad51*); **c**, RY580T1 (*rad54*), RY582T1 (*rad55*) and RY581T1 (*rad57*). The probe for the Southern blots is indicated in **d**. Lanes 0 show the *MAT_a* and *MAT_{distal}* *StyI* fragments. The *HO* cut fragment appears at 0.5 h after *HO* induction, followed by the appearance of *MAT_α* product beginning at 1.5 h in the *Rad⁺* strain. The *MAT_α* product band cannot be observed in *rad51*, *rad54*, *rad55* or *rad57* strains. Arrows indicate the formation of single-stranded (ss) DNA tails that result from the inability of *StyI* to cleave the ssDNA which has extended beyond one or more distal *StyI* sites⁴. **d**, A molecular model of *MAT* switching. *HO* endonuclease induces a double strand cut within the *MAT* locus at the Y/Z junction. Exonucleolytic degradation to the right of the DSB creates a 3' single-stranded tail that invades the homologous sequence of the silent copy donor *HML_αZ* region. Strand invasion and DNA extension can be monitored by a PCR assay using primers in *HML_α* and distal to *MAT_α* (labelled PCR_α and PCR_β). When strand extension reaches the X region of *HML_α*, the displaced strand anneals to the X sequence from *MAT_a* and serves as a template for the replacement of the *Y_a* sequence with *Y_a* sequence. Formation of the new *MAT_α/Y_a* junction can be monitored using PCR primers located in the unique *MAT* proximal sequence (PCR_γ) and in the *Y_a* sequence (PCR_δ).

METHODS. Cells carrying a galactose-inducible *HO* gene^{17,18} were induced for *HO* as described⁴. Briefly, cells were grown in liquid YEPLactate medium overnight to a density of 10^7 cells per ml and then induced to express the *HO* endonuclease by the addition of 1/10 vol of 20% galactose. After 30 min, 1/10 vol of 20% glucose was added to downregulate the expression of *GAL::HO*. DNA was extracted at the given time points by disrupting the cells with glass beads and phenol. The DNA was digested with *StyI* and electrophoresed in an alkaline denaturing gel. The Southern blot was probed with a *MAT*-distal fragment. Deletions of different *RAD* genes in strain R166 was accomplished by one-step gene replacement¹⁹. The deletions used included: *rad51::URA3* (ref. 8) and *rad51::LEU2* from plasmid pAM50 (ref. 20), *rad52::LEU2* from plasmid pSM20 (ref. 21), *rad54::URA3* derived from plasmid pLS101 (ref. 22), *rad55::LEU2* from plasmid pSTL11 (ref. 23) and plasmid *rad57::LEU2* from plasmid pSM51 (ref. 21).

a centromere-containing plasmid, pNSU178. As shown in Fig. 2b, *MAT* switching failed to occur in a *rad51* strain, although a *Rad⁺* strain showed efficient switching. In plasmid pNSU216 (Fig. 2d), we replaced the silent *HMR* region with an expressed *MAT_α-inc* sequence that cannot be cleaved by *HO* but serves as an efficient donor. *GAL::HO*-induced recombination was as efficient in a *rad51* strain as in *Rad⁺* (Fig. 2e, f). Thus *MAT_α* can recombine with an unsilenced donor in the absence of *RAD51*. Similar results were obtained with *rad54*, *rad55* and *rad57* strains carrying pNSU216; however, *rad52* still prevented switching (data not shown).

As another example of *HO*-induced recombination between unsilenced sequences, we examined a plasmid containing two copies of the *E. coli lacZ* gene, one of which contains an *HO* cut site¹⁰. Recombination required *RAD52* but not *RAD51*, *RAD54*, *RAD55* or *RAD57* (Fig. 3 and data not shown).

In strains carrying pNSU216, the chromosomal *MAT_a* locus was also cut by *HO*. Approximately 30% of the *rad51* cells survived and switched to *MAT_α-inc*. The ability of the plasmid-borne *MAT_α-inc* to act as a donor for the chromosomal *MAT_a* gene is evident in a Southern blot (Fig. 2e). Thus a chromosomal *MAT_a* locus can also switch to *MAT_α-inc* if the donor is not silenced and carried on a plasmid.

Although *MAT_α-inc* sequences on a plasmid were able to serve as donors in *rad51* strains, this was not the case when the same sequences were located on a chromosome. This has been shown in three ways. First, a *MAT_a/MAT_α-inc rad51/rad51*

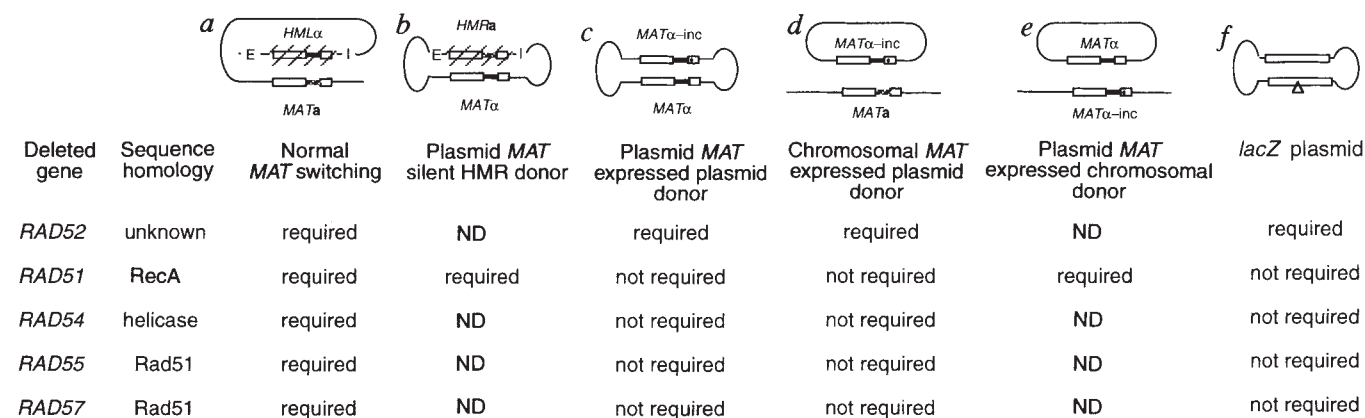


FIG. 3 Requirements for *RAD* genes in HO-induced recombination. The effect of different *RAD* gene deletions on HO-induced recombination is summarized for six different assays. *a*, A chromosomal *MATa* locus switches using the silent copy donor, *HMLα* located 200 kb away. *b*, *MATa* and the silent *HMRα* locus are carried on a centromeric plasmid pNSU178 (Fig. 2a). *c*, *MATa* on plasmid pNSU216 (Fig. 2d) is repaired by an expressed *MATα-inc* donor on the same plasmid. *d*, *MATa* on the chromosome is repaired by *MATα-inc* on plasmid pNSU216 (the plasmid *MATa* is not shown). *e*, An expressed *MATα-inc* on the chromosome is used as a donor to switch *MATa* on plasmid pNSU215. *f*, HO

induced recombination between inverted *lacZ* sequences in plasmid pJF5 (ref. 10). The HO cut site is indicated by a triangle. Possible functional roles of the *RAD* genes are indicated. No biochemical role has yet been assigned to the *Rad52* protein, which interacts physically with *Rad51* (refs 24, 25). *Rad51* has significant similarity to the *RecA* protein of *Escherichia coli* and other bacteria^{8,20,25-27}. *Rad57*, and to a lesser extent *Rad55*, exhibits homology with *Rad51* (and thus with *RecA*), but no biochemical studies have yet been reported^{28,29}. *Rad54* shares consensus protein sequence motifs with many DNA helicases³⁰. ND, not done.

diploid fails to repair an HO-induced break, whereas *Rad*⁺ strains readily become *MATa-inc/MATa-inc* (data not shown). Second, a *rad51 MATa* haploid failed to switch when the chromosomal *HMLα* locus was replaced with an expressed *MATa-inc* gene to act as donor (data not shown). Finally, a *rad51 MATa-inc* strain was transformed with pNSU215 which is identical to pNSU216 (Fig. 2d), except that it does not carry the fragment containing the *MATa-inc* donor. Thus, this plasmid can only be repaired by recombining with a donor carried on the chromosome. The chromosomal *MATa-inc* gene efficiently donated sequences to repair *MATa* on plasmid pNSU215 in *Rad*⁺ strains but not in a *rad51* strain (data not shown). The idea that DNA sequences on a plasmid are functionally different from those on a chromosome has previously been documented for chromatin structure¹¹, the expression of *HML* and *HMR*^{12,13}

or the function of ARS sequences¹⁴. Our results suggest that there are also differences in chromatin structure that significantly affect recombination. A *MATa-inc* gene located on a chromosome is dependent on *RAD51* to act as a donor, whereas the same gene located on a plasmid is *RAD51*-independent, even though *MATa-inc* is expressed on the chromosome as well as on a plasmid. The use of a silent *HMR* gene as a donor always requires *RAD51*.

Moreover, different states of DNA have different requirements for DNA repair; the removal of ultraviolet-induced damage from *HMLα* requires *RAD7* and *RAD16*, but the repair of *MAT* (or an unsilenced *HML*) does not^{15,16}. We suggest that *RAD51*, *RAD54*, *RAD55* and *RAD57* are not needed for recombination *per se*, but are required to gain access to otherwise inaccessible regions of chromatin. □

Received 30 August; accepted 10 November 1994.

- Haber, J. E. *Trends Genet.* **8**, 446-452 (1992).
- Strathern, J. N. in *Genetic Recombination* (eds Kucherlapati, R. & Smith, G. R.) 445-464 (Am. Soc. Microbiol., Washington DC, 1989).
- Raveh, D., Hughes, S. H., Shafer, B. K. & Strathern, J. N. *Molec. gen. Genet.* **220**, 33-42 (1989).
- White, C. I. & Haber, J. E. *EMBO J.* **9**, 663-674 (1990).
- Haber, J. E., Ray, B. L., Kolb, J. M. & White, C. I. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3363-3367 (1993).
- Ivanov, E. L., Sugawara, N., White, C. I., Fabre, F. & Haber, J. E. *Molec. cell. Biol.* **14**, 3414-3425 (1994).
- Malone, R. E. & Esposito, R. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 503-507 (1980).
- Aboussekhr, A., Charet, R., Adjiri, A. & Fabre, F. *Molec. cell. Biol.* **12**, 3224-3234 (1992).
- Petes, T. D., Malone, R. E. & Symington, L. S. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* (eds Broach, J. R., Pringle, J. & Jones, E. W.) 407-521 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1991).
- Fishman-Lobell, J., Rudin, N. & Haber, J. E. *Molec. cell. Biol.* **12**, 1292-1303 (1992).
- Simpson, R. T. *Progr. Nucleic Acids Res. molec. Biol.* **40**, 1896-1899 (1993).
- Brand, A. H., Breiden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. *Cell* **41**-48 (1985).
- Mahoney, D. J. & Broach, J. R. *Molec. cell. Biol.* **9**, 4621-4630 (1989).
- Collins, I. & Newlon, C. S. *Molec. cell. Biol.* **14**, 3524-3534 (1994).
- Bang, D. D., Verhage, R., Goosen, N., Brouwer, J. & van de Putte, P. *Nucleic Acids Res.* **20**, 3925-3931 (1992).

- Verhage, R. et al. *Molec. cell. Biol.* **14**, 6135-6142 (1994).
- Jensen, R. & Herskowitz, I. *Cold Spring Harbor Symp. quant. Biol.* **49**, 97-104 (1984).
- Nickoloff, J. A., Singer, J. D., Hoekstra, M. F. & Heffron, F. J. *molec. Biol.* **207**, 527-541 (1989).
- Rothstein, R. J. *Meth. Enzym.* **202**-211 (1983).
- Basile, G., Aker, M. & Mortimer, R. K. *Molec. cell. Biol.* **12**, 3235-3246 (1992).
- Schild, D., Calderon, I. L., Contopoulou, C. R. & Mortimer, R. K. in (eds Friedberg, E. C. & Bridges, B. A.) *Cellular Responses to DNA Damage* 417-427 (Liss, New York, 1983).
- Rattray, A. & Symington, L. *Genetics* (in the press).
- Lovett, S. T. & Mortimer, R. K. *Genetics* **116**, 547-553 (1987).
- Milne, G. T. & Weaver, D. T. *Genes Dev.* **7**, 1755-1765 (1993).
- Shinohara, A., Ogawa, H. & Ogawa, T. *Cell* **69**, 457-470 (1992).
- Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. *Science* **259**, 1896-1899 (1993).
- Sung, P. *Science* **265**, 1241-1243 (1994).
- Kans, J. A. & Mortimer, R. K. *Gene* **105**, 139-140 (1991).
- Lovett, S. T. *Gene* **142**, 103-106 (1994).
- Emery, H. S., Schild, D., Kellogg, D. E. & Mortimer, R. K. *Gene* **104**, 103-106 (1991).

ACKNOWLEDGEMENTS. Supported by a grant from the NIH: J.F.-L. was a trainee of a USPHS Training Grant in Genetics. Plasmid pHF1 was kindly provided by H. Friesen and P. Sadowski. We thank S. Lovett, D. Schild and L. Symington for other *RAD* gene deletions; J. Kolb for PCR amplifications and for technical assistance; and S. Lovett, M. Lichten and S. Rosenberg and members of our laboratory for comments on the manuscript.

Nature's sister journals in review

As a service to those readers of *Nature* who do not see its sister journals regularly, the following pages refer to some of the content of *Nature Genetics*, *Nature Structural Biology* and — published for the first time this week — *Nature Medicine*.

Cellular immunity to HIV

AN explanation of some people's apparent immunity to infection by HIV is suggested by a study of the immune status of a group of prostitutes in The Gambia in West Africa, who carry in their blood cytotoxic lymphocytes (CTLs) specifically active against identified immunologically active peptides (epitopes) from the HIV proteins. The importance of this finding is that it may point to better routes to the design of vaccines against HIV than those at present being followed.

The international study, published in the first issue of *Nature Medicine*, involves groups at the John Radcliffe Hospital, Oxford, the Institute of Cancer Research in London, the Institute of Medical Science at the University of Tokyo and the UK Medical Research Council Laboratories at Fajara in The Gambia (S. Rowland-Jones *et al. Nature Med.* 1, 59–64; 1995).

The study involved 20 women, six of whom were apparently free from infection by either HIV-1 or HIV-2 (which has given way to HIV-1 as the predominant strain in The Gambia) despite long-standing exposure to HIV. While CTLs specific for HIV peptides are naturally evident in the blood of infected people, their presence in those without overt infection (represented by the absence of antibodies, for example) has no natural explanation except prior infection.

Of the six high-risk women free from the disease, CTLs specifically active against HIV peptides were evoked in only three. The likely inference, the authors say, is that these three women were at some stage exposed to possibly low-level

infection, but that infected cells were eliminated by CTLs before the infection became established. It is not yet known how permanent this response will be, although there is every likelihood that clonal memory will remain in the thymus for some time.

There is no explanation of the apparent immunity of the other three high-risk women in the study — although they may have a similar response to other HIV peptides than the two used in The Gambia. Or they may simply have been lucky.

Sex and skeletal genes

THE ramifications of sex determination in mammals continue to proliferate. It has been known for nearly four years that the masculinity of normal human males derives from their Y chromosomes and in particular from the gene called *SRY*, first characterized (by its analogue *Sry* in the mouse) by Goodfellow and colleagues in an elegant experiment with transgenic mice (*Nature* 351, 117–121; 1991).

But *Sry* is also one of a family of DNA-regulatory genes called *Sox*, one of which (*Sox9*) resembles in structure a region of human chromosome 17, mutations of which are responsible for two known autosomal genetic effects — sex reversal (by means of which genetic males are phenotypically female) and the rare but often fatal skeletal disease called campomelic dysplasia (CD). Two-thirds of the XY males afflicted by CD are phenotypically female.

Now, in the current issue of *Nature Genetics*, Koopman's group at the University of Queensland, Brisbane, have described the expression of the regular version of *Sox9* during embryogenesis in the mouse, finding that the gene is primarily expressed before and during the formation of cartilage in the regions in which skeletal bone will eventually be laid down (E. Wright *et al. Nature Genet.* 9, 15; 1995).

The regulatory character of *Sox9* is suggested by its disappearance from embryonic tissues after the formation of cartilage (and certainly at 13.5 days post-conception) as well as by the presence near the upstream end of the gene of a nucleotide sequence characteristic of an amino-acid sequence known to bind to DNA (the 'MGM box'). *Sox9* also transiently appears in the tubular structures of

the heart, the spinal cord and the notochord.

The mouse gene *Sox9* is located on chromosome 11, and close enough on the mouse map to the mutation *tail-short* to suggest that the genes responsible are the same. That provides a potentially valuable mouse model for a crucial genetic step in the development of the mammalian skeleton as well as for the gene responsible for CD in people. Disappointingly perhaps, there is no sign of sex reversal in XY male mice heterozygous for *tail-short*. (Homozygotes do not survive.)

Transgenic Alzheimer's

IN the search for an animal model of Alzheimer's disease (AD), the transgenic mouse is an obvious route; give a mouse a suitable transgene under the control of a suitable promoter, and you may find that the mice develop a condition comparable with AD. But what gene, and what promoter?

A group from the Jerome H. Holland Laboratory at Rockville, Maryland, has now developed three independent strains of transgenic mice designed to generate the mouse version of β -amyloid protein (otherwise A β) directly within their neurons (F. M. LaFerla *et al. Nature Genet.* 9, 21–30; 1995). Earlier transgenic mice carrying the gene for the much larger molecule called amyloid precursor protein (APP), either human or that found in the mouse, have not successfully mimicked AD.

The trick is to drive the gene corresponding to the 42 amino acids of the mouse version of A β by the promoter region of the neurofilament-light gene (*NF-L*), which is active throughout the adult life of the mouse, but almost exclusively in neurons. The characteristic extracellular plaques formed in the brains of human patients with AD are made predominantly from (human) A β , but they form outside the cell and may yet be a consequence, rather than a cause, of the underlying condition.

LaFerla *et al.* now report that hybridization with a nucleotide probe and immuno-staining with antibodies against fragments of A β show that the transgene is actively expressed in the hippocampus and cerebral cortex of the transgenic mice, "regions significantly compromised in AD". They also say that the transgenic mice are more prone to early death than

**nature
medicine**

Nature Medicine — other points

The centrosome of the fertilized human egg, by which the haploid nuclei of the germ cells are drawn together into a single nucleus, is provided by the male — which may be relevant in infertility (pages 47–52).

Cationic liposomes carrying DNA on their outer surfaces can effectively transport genes into cells, but gene therapy trials with cystic fibrosis genes should that nuclear incorporation is ineffective (pages 39–46). □

**nature
genetics**

normal controls; for the most actively expressing of the three strains, half the mice were dead at 12 months, compared with the normal life expectancy of 24 months.

The authors report cell death in these regions by both necrosis and apoptosis, and say that dissection of the second mechanism may suggest means of therapy of human AD. But others are not so sure. In a comment on the report (*Nature Genet.* 9, 4-7; 1995), Bruce T. Lamb of the Johns Hopkins University School of Medicine in Baltimore, Maryland, notes that it has recently been found that mutations of the APP gene, which codes for a transmembrane protein whose physiological function is unknown, lead to over-production of A β . He agrees with LaFerla *et al.* that transgenic mice carrying mutations of that kind may be needed to disentangle the pathogenesis of AD.

Do mice make rats?

DOES *Rattus*, the rat, deserve a genome project of its own, distinct from that of the mouse? An affirmative answer has been given by a group of 20 geneticists organized around Dr Eric Lander of the Whitehead Institute in Cambridge, Massachusetts, which has produced a first linkage map of the 21 chromosomes of the rat genome based on 432 markers of sites of simple sequence length polymorphisms (SSLPs). In the modern idiom, the account of this first step towards a physical map of the rat genome (H. J. Jacob *et al.* *Nature Genet.* 9, 63-69; 1995) explains how enquirers can gain access to the data through a server located at Massachusetts Institute of Technology.

As it stands, the linkage map of the rat is only coarse, with fewer than a tenth as many markers as in the current editions of the mouse and human maps. But, with the exception of chromosomes 9 and 16, where there may be missing sections, the authors believe their markers are more or less uniformly distributed across the genome.

The data have been gathered by groups based at the Massachusetts General Hospital, the University of California at San Francisco, the University of Wisconsin, the Free University of Brussels, the University of Göteborg, the Karolinska Institute (Stockholm) and the Charles University, Prague. The SSLPs used in the mapping were determined in a first-generation cross of the brown Norwegian strain of *Rattus norvegicus* with another (SHR) bred for hypertension studies.

Describing their linkage map as a "milestone on the path to realizing the full potential of rat genetics", Jacob *et al.* foresee a repetition for the rat of the genome projects now under way for mice and men.

But Wayne N. Frankel from the Jackson Laboratories in Bar Harbor, Maine, in a comment on the linkage map, raises the interesting question whether, given the close genetic relationship between rats and mice, it may be possible (and more economical) to identify and place on a physical map the genes of the rat by reference to their homologues in the mouse.

Atherosclerosis by MRI

TELLING which of a person's arteries are occluded by atherosclerotic plaques, and where, is at present an uncertain business involving the invasive techniques of angiography or angiography. But, in the first issue of *Nature Medicine*, Michael P. Skinner and colleagues from the University of Washington at Seattle and the Westmead Hospital in New South Wales, Australia, have been able to show that new developments in MRI (magnetic resonance imaging) can provide more useful information as well as a better understanding of the nature of the lesions (M. P. Skinner *et al.* *Nature Med.* 1, 69-73; 1995).

The authors describe a series of imaging studies with five of six live rabbits fed from birth with a diet including cholesterol and peanut oil, and whose abdominal aortas had been damaged soon after birth by pulling an inflated balloon through them to promote formation of atherosclerotic plaques. By using a technique ('time-of-flight' MRI) in which the magnetic field is provided by coils suitably placed around the abdomens of the rabbits, the authors are able to use the time resolution of the detection equipment to provide what are apparently cross-sectional images of body tissues.

One of the most vivid lessons of the images accompanying the published paper is that of the progressive thickening of the walls of a rabbit artery in the course of half a year, and the concomitant reduction of the cross-sectional area of the artery (the lumen) by a factor of nearly three.

But there is more to atherosclerosis and its risk than the progressive reduction of the column available for blood flow in the arteries. The MRI images have also been able to capture some occasions on which atherosclerotic plaques are in the process of detaching themselves from the arterial walls.

Researchers will welcome these developments as a means of throwing further light on the mechanism of the development of atherosclerosis, known to involve the interaction of several lymphokines and growth factors with the cells of the arterial wall and its substructures. (See the review by Russell Ross, one of the authors of this article, in *Nature* 362, 801-809; 1993.) Among other things, there will be great interest in knowing why therapy for

atherosclerosis is often more effective than the mechanical shape of the arteries revealed by angiography would suggest it should have been.

Skinner *et al.* say that the application of this version of MRI to human patients will be complicated by the greater mass of people (than rabbits), which will reduce the strength of the MRI signal, as well as by the need to investigate the conditions of the arteries of the heart, which is in motion. They refer to the possibility of using the pulsation of the heart to trigger MRI pulses. At the outset, when applied to people, the technique is most likely to be used to assess the effectiveness of therapy, but the possibility that the system may be used for surveys of the condition of the arteries as a whole is not ruled out.

Electron diffraction in ice

ONE of the most sophisticated uses yet of the technique called electron cryomicroscopy is that for the determination of the molecular structure of the skeletal muscle calcium release channel now reported by a group from the Baylor College of Medicine in Houston, Texas, with collaboration by the Fritz Haber Institute in Berlin (I. I. Serysheva *et al.* *Nature Struct. Biol.* 2, 18-23; 1995).

The protein molecules that form the channel have an estimated relative molecular mass of 2.4×10^6 . The channel complex has the general shape of a square with 270 Å sides, but with an open 'pretzel-like' appearance. The bulk of the channel is 125 Å thick, but there is a loop-like structure projecting 65 Å on the presumed lumen side of the complex which is supposed also to carry the domain that spans the sarcoplasmic reticulum.

Electron cryomicroscopy is based on the idea that protein molecules may be embedded in glassy ice without changing shape. Although the protein structures will, in general, be oriented randomly relative to each other, each will yield a separate diffraction image. By comparison of these images, supposedly related to each other by random rotations of an identical protein structure, it is then possible to infer the distribution of electron-scattering density in the structure as a whole.

Serysheva *et al.* have analysed a total of 3,000 separate electron-diffraction patterns from separate calcium-releasing channels taken, as it happens, from fast-twitch skeletal muscle in the rabbit. They say that their study is the first in which the "angular reconstitution" approach has been used to solve a molecular structure. Because the channels used in this study are inactive and therefore closed to the transport of calcium, the absence of a central hole through the complex is not surprising. □